TRANSACTIONS
OF THE ASME
Series B

Journal of ENGINEERING FOR INDUSTRY

November 1960

Divisions: AVIATION • MACHINE DESIGN • MANAGEMENT • PROCESS INDUSTRIES • MAINTENANCE AND PLANT ENGINEERING • RAILROAD • PRODUCTION ENGINEERING • MATERIALS HANDLING • RUBBER AND PLASTICS • PETROLEUM • SAFETY • TEXTILE • WOOD INDUSTRIES • **Research Committees:** METAL PROCESSING • FLOW OF BULK MATERIALS • ELEVATORS • MECHANICAL SPRINGS

Published Quarterly by
THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS

TRANSACTIONS OF THE ASME
Series B

Journal of
ENGINEERING FOR INDUSTRY

VOLUME 82 • SERIES B • NUMBER 4

November 1960

CONTENTS

283 **Tensile and Short-Time Creep Properties of N-155 Alloy Sheet** (59—A-27)
E. C. Bernett

303 **A Photoelastic Analysis of Machining Stresses** (59—A-37)
E. Usui and H. Takeyama

309 **Correlation of Bendability of Materials With Their Tensile Properties** (59—A-110)
J. Datsko and C. T. Yang

315 **Comparisons Between the Shearing Properties of Alpha-Brass as Derived From the Cutting Process and From Static and Impact Torsion Tests** (59—A-38)
S. S. Chang and W. B. Heginbotham

324 **The Role of Friction in Metal Cutting** (59—A-132)
S. Kobayashi and E. G. Thomsen

333 **A Critical Comparison of Metal-Cutting Theories With New Experimental Data** (59—A-131)
S. Kobayashi, R. P. Herzog, D. M. Eggleston, and E. G. Thomsen

348 **New Developments in the Theory of the Metal-Cutting Process, Part I** (59—A-243)
P. Albrecht

359 **Structural Weight Approximation for a Bell-Nozzle Divergent Section** (59—A-69)
G. D. Brewer and E. Levin

363 **Optimum Tolerances of Sheet Materials for Flight Vehicles** (59—A-70)
G. A. Hoffman

369 **Mechanical Vibration Filters With Nonlinear Characteristics** (59—A-42)
W. J. Worley

377 **Refinement of Finite Difference Calculations in Kinematic Analysis** (59—A-53)
B. W. Shaffer and I. Krause

382 **The Measurement of Very Low Frictional Torques in Rotating Equipment** (59—A-43)
M. Zaid and I. S. Tolins

387 **Design by Quantitative Factor of Safety** (59—A-48)
Hsuan-Loh Su

393 **Study of Critical Velocity of Stick-Slip Sliding** (59—A-146)
B. R. Singh

399 **Inertial Effects in a Multiple-Ball Transmission** (59—A-14)
W. S. Rouverol

407 **Some Applications of the Cycloid in Machine Design** (59—A-134)
E. P. Pollitt

415 **A Discussion of the Vibration Characteristics of a Simple Mechanical Connection** (59—A-41)
Arthur Sorensen, Jr.

423 **Deformation and Moments in Elastically Restrained Circular Plates Under Arbitrary Load or Linear Thermal Gradient** (59—A-39)
M. Zaid and M. Forray

439 **Dimensionless Parameters for Helical Compression Springs** (59—A-30)
R. J. Erisman

444 **Errata on previously pubished paper by** *R. C. Brewer*

EDITORIAL DEPARTMENT

Editor, J. J. JAKLITSCH, JR.
Consulting Editor, LEO BLODGETT
Production, JEAN NORTH
Art Editor, HENRY BALFOUR

OFFICERS OF THE ASME

President, W. L. CISLER
Secretary, O. B. SCHIER, II
Treasurer, E. J. KATES
Assistant Treasurer, H. J. BAUER

PUBLICATIONS COMMITTEE

Chairman, B. G. A. SKROTZKI
HENDLEY N. BLACKMON
R. D. MINDLIN
MARTIN GOLAND
VITO L. SALERNO
Junior Representatives
A. S. GOLDSTEIN
J. W. FOLLANSBEE

REGIONAL ADVISORY BOARD

Region I, H. M. ROGERS, JR.
Region II, GLENN R. FRYLING
Region III, F. J. HEINZE
Region IV, FRANCIS C. SMITH
Region V, H. M. CATHER
Region VI, C. R. EARLE
Region VII, M. S. UMBENHAUER
Region VIII, CARL J. ECKHARDT

Published quarterly by The American Society of Mechanical Engineers. Publication office at 20th and Northampton Streets, Easton, Pa. The editorial department is located at the headquarters of the Society, 29 West Thirty-Ninth Street, New York 18, N. Y. Cable address, "Mechaneer," New York. Price $3.00 a copy, $10.00 annually; to members, $1.50 a copy, $5.00 annually. Add 50 cents for postage to all countries outside the United States and Canada. Changes of address must be received at Society headquarters seven weeks before they are to be effective on the mailing list. Please send old as well as new address. . . . By-Law: The Society shall not be responsible for statements or opinions advanced in papers or . . . printed in its publications (B13, Par. 4). . . . Second-class mail privileges authorized at Easton, Penna. . . . This publication is authorized to be mailed at the special rates of postage prescribed by Section 132.122 . . . © Copyrighted, 1960, by The American Society of Mechanical Engineers. Reprints from this publication may be made on condition that full credit be given the Transactions of the ASME, Series B—Journal of Engineering for Industry, and the author, and that date of publication be stated.

TRANSACTIONS OF THE ASME
is indexed by the Engineering Index, Inc.

TRANSACTIONS OF THE ASME Series C

Journal of
HEAT TRANSFER
Vol. 82 • Series C • No. 4

CONTENTS FOR NOVEMBER, 1960

265 A Variable Density Single-Fluid Model for Two-Phase Flow With Particular Reference to Steam-Water Flow, by S. G. Bankoff. (59—HT-7)

273 Transient Temperatures and Thermal Stresses in Hollow Cylinders Due to Heat Generation, by John E. Schmidt and George Sonnemann. (59—A-144)

279 Thermal-Contact Resistance in Finned Tubing, by Karl A. Gardner and T. C. Carnavos. (59—A-135)

294 Mass Transfer, Flow, and Heat Transfer About a Rotating Disk, by E. M. Sparrow and J. L. Gregg. (58—A-107)

303 Heat Transfer and Effectiveness for a Turbulent Boundary Layer With Tangential Fluid Injection, by R. A. Seban. (59—A-177)

313 Sublimation Mass Transfer Through Compressible Boundary Layers on a Flat Plate, by T. K. Sherwood and Olev Trass. (59—A-137)

325 Laminar Skin Friction and Heat Transfer on Flat Plates With Wedge-Shaped Grooves in Flow Direction, by T. F. Irvine, Jr., and E. R. G. Eckert. (59—A-143)

333 Heat Transfer to Liquid Metals With Variable Properties, by R. Viskanta and Y. S. Touloukian. (59—A-148)

341 A Summary of Experiments on Turbulent Heat Transfer From a Nonisothermal Flat Plate, by W. C. Reynolds, W. M. Kays, and S. J. Kline. (59—A-157)

349 An Experimental Study of the Effects of Nonuniform Wall Temperature on Heat Transfer in Laminar and Turbulent Axisymmetric Flow Along a Cylinder, by R. Eichhorn, E. R. G. Eckert, and A. D. Anderson. (59—A-151)

360 Geometric Factors for Radiative Heat Transfer Through an Absorbing Medium in Cartesian Coordinates, by A. K. Oppenheim and J. T. Bevans. (59—A-206)

369 Thermal Radiation From a Cylindrical Enclosure With Specified Wall Heat Flux, by C. M. Usiskin and R. Siegel. (59—A-159)

375 Application of Variational Methods to Radiation Heat-Transfer Calculations, by E. M. Sparrow. (59—A-120)

381 Certain Thermodynamic and Transport Properties of the Atmospheres of Venus, Mars, and Jupiter, by W. E. Ibele and T. F. Irvine, Jr. (59—A-102)

TRANSACTIONS OF THE ASME Series D

Journal of
BASIC ENGINEERING
Vol. 82 • Series D • No. 3

CONTENTS FOR SEPTEMBER, 1960

505 Theoretical and Experimental Analysis of Hydrostatic Thrust Bearings, by R. C. Elweil and B. Sternlicht. (59—A-152)

513 Precise Measurement of Clothes Insulation During Controlled Operation of the Human, by A. S. Iberall. (59—A-197)

528 Radial Distributions of Temporal-Mean Peripheral Velocity and Pressure for Fully Developed Turbulent Flow in Curved Channels, by A. W. Marris. (59—A-51)

539 Flow Phenomena of Partially Enclosed Rotating Disks, by L. A. Maroti, G. Deak, and F. Kreith. (59—A-49)

553 Roughness Effects on Frictional Resistance of Enclosed Rotating Disks, by R. E. Nece and J. W. Daily. (59—A-55)

563 Rotating Wakes in Vaneless Diffusers, by R. C. Dean, Jr., and Y. Senoo. (59—A-104)

575 Incidence and Deviation-Angle Correlations for Compressor Cascades, by Seymour Lieblein. (59—A-171)

588 A Study of Boundary-Layer Characteristics of Turbomachine Blade Rows and Their Relation to Over-All Blade Loss, by W. L. Stewart, W. J. Whitney, and R. Y. Wong. (59—A-23)

593 Laminar Boundary Layers in Oscillatory Flow, by P. G. Hill and A. H. Stenning. (59—A-265)

609 Experimental Determination of Statistical Properties of Two-Phase Turbulent Motion, by S. L. Soo, H. K. Ihrig, Jr., and A. F. El Kouh. (59—A-59)

622 The Turbulent Boundary Layer at a Plane of Symmetry in a Three-Dimensional Flow, by J. P. Johnston. (59—A-72)

629 Heat-Transfer Characteristics of Hot-Film Sensing Element Used in Flow Measurement, by S. C. Ling. (59—A-172)

635 The Adiabatic Bulk Modulus of Normal Paraffin Hydrocarbons From Hexane to Hexadecane, by R. E. Rolling and C. J. Vogt. (59—A-169)

645 Variable Flow Resistance With Adjustable Multihole Orifice Plates in Series, by J. Sherman, F. A. Grochowski, and J. E. Sharbaugh. (59—A-182)

654 Cycle-Dependent Stress Relaxation of A-286 Alloy, by A. S. Ross and JoDean Morrow. (59—A-116)

661 Low-Cycle Fatigue of Two Nickel-Base Alloys by Thermal-Stress Cycling, by F. J. Mehringer and R. P. Felgar. (59—A-58)

671 The Stability of Metals Under Cyclic Plastic Strain, by L. F. Coffin, Jr. (59—A-100)

683 Roll-Force and Torque Coefficients for Hot-Strip Steel Mill, by B. N. Garudachar and H. A. Peterson. (59—A-153)

689 Steady-State Creep of a Thick-Walled Cylinder Under Combined Axial Load and Internal Pressure, by Iain Finnie. (59—A-57)

695 Burst Strengths and Deformations of Welded Composite Turbine Wheels as Related to Weld Quality and Plasticity Calculations, by A. G. Holms and A. J. Repko. (59—A-28)

710 The Twin Throat Venturi: A New Fluid-Flow Measuring Device, by A. A. Kalinske. (59—A-154)

717 A Study of Viscosity Effect and Its Compensation on Turbine-Type Flowmeters, by W. F. Z. Lee and Henning Karlby. (59—A-105)

729 The Quadrant Edge Orifice—A Fluid Meter for Low Reynolds Numbers, by M. Bogema and P. L. Monkmeyer. (59—A-140)

735 Small-Diameter-Orifice Metering, by T. J. Filban and W. A. Griffin. (59—A-101)

741 Analysis of Axially Loaded Annual Shells With Applications to Welded Bellows, by M. Hetényi and R. J. Timms (59—A-175)

TRANSACTIONS OF THE ASME Series A

Journal of
ENGINEERING FOR POWER

Volume 82 • Series A • Number 4

CONTENTS FOR OCTOBER, 1960

247 Throat Tap Nozzles Used for Accurate Flow Measurements, by K. C. Cotton and J. C. Westcott. (59—A-174)

264 Water-Treatment, Corrosion, and Internal-Deposit Studies for Eddystone, by R. C. Ulmer, H. A. Grabowski, and R. C. Patterson. (59—A-147)

275 The Selection of Materials and Fabricating Techniques for the Eddystone Boiler and Sulzer Control Valves, by E. C. Chapman and R. E. Lorentz, Jr. (59—A-318)

286 Development Associated With the Superpressure Turbine for Eddystone Station Unit No. 1, by W. E. Trumpler, Jr., A. F. LeBreton, E. A. Fox, and R. B. Williamson. (59—A-288)

293 Material Selection and Fabrication, Main Steam Piping for Eddystone No. 1, 1200-F and 5000-Psi Service, by R. H. Caughey and W. G. Benz, Jr. (59—A-327)

315 Solar Turbo Power Plant Design, by D. B. Mackay and E. L. Leventhal. (59—A-74)

325 Analysis of Solar-Furnace Performance in Mechanical Testing at Extremely High Temperatures, by G. S. Leon and M. E. Shank. (59—A-79)

TRANSACTIONS OF THE ASME Series E

Journal of
APPLIED MECHANICS

Volume 82 • Series E • No. 3

CONTENTS FOR SEPTEMBER, 1960

TECHNICAL PAPERS

381 On Three and Two-Dimensional Disturbances of Pipe Flow, Kurt Spielberg and Hans Timan. (60—APM-12)

390 On the Indeterminateness of the Boundary Conditions for the Mixing of Two Parallel Streams, K. T. Yen. (59—A-68)

393 Thermal Transfer in Turbulent Gas Streams. Effect of Turbulence on Local Transport From Spheres, W. W. Thort, R. A. S. Brown, and B. H. Sage. (60—APM-30)

403 Hydrodynamic Entrance Lengths for Incompressible Laminar Flow in Rectangular Ducts, L. S. Han. (59—A-82)

410 Stability of a Turbine-Generator Rotor Including the Effects of Certain Types of Steam and Bearing Excitations, A. H. Landzberg. (60—APM-9)

417 On Saint Venant's Principle: Elastic Shells and Plates, P. M. Naghdi. (60—APM-11)

423 On the Airy Stress Functions of Plane Dislocation Problems, F. D. Ju. (60—APM-14)

429 The Effect of Pressure on the Bending Characteristics of an Actuator System, Paul Seide. (60—APM-16)

438 The Contact Problem for Viscoelastic Bodies, E. H. Lee and J. R. M. Radok. (59—A-83)

445 Stresses in Curved Beams of Noncircular Center Line, M. M. Abbassi. (60—APM-21)

455 Shear Buckling of Bars, G. Herrmann and A. E. Armenakas. (59—A-91)

458 Buckling of Circular Cones Under Axial Compression, Leslie Lackman and Joseph Penzien. (60—APM-17)

461 The Additional Deflection of a Cantilever Due to the Elasticity of the Support, W. J. O'Donnell. (60—APMW-3)

465 Thermoelastic Stresses in Beams, E. S. Barrekette. (60—APM-28)

474 Elastic Instability of Rectangular Sandwich Panel of Orthotropic Core With Different Face Thicknesses and Materials, C. C. Chang and I. K. Ebcioglu. (60—APM-26)

481 Transient and Residual Thermal Stresses in an Elastic-Plastic Cylinder, H. G. Landau and E. E. Zwicky, Jr. (60—APM-4)

489 On Strain-Hardened Circular Cylindrical Shells, Nicholas Perrone and P. G. Hodge, Jr. (60—APM-20)

496 The Elastoplastic Thick-Walled Sphere Subjected to a Radial Temperature Gradient, G. R. Cowper. (60—APM-22)

501 Rupture Instability of Membranes and Solids, S. F. Borg. (60—APM-1)

505 Expansion of Thick-Walled Cylinders Fabricated From Cold Bent Plates, Irwin Berman. (60—APMW-2)

513 Dispersion of Flexural Waves in an Elastic, Circular Cylinder, Yih-Hsing Pao and R. D. Mindlin. (59—A-84)

521 Modeling of Large Transient Elastic and Plastic Deformations of Structures Subjected to Blast Loading, W. E. Baker. (60—APM-18)

528 Shock Spectra of a Nonlinear System, W. T. Thomson. (60—APM-23)

535 Forced Flexural Vibrations of Sandwich Plates in Plane Strain, Yi-Yuan Yu.

541 Extensional Vibrations and Waves in a Circular Disk and a Semi-Infinite Plate, D. C. Gazis and R. D. Mindlin. (60—APM-3)

548 Effect of Small Hub-Radius Change on Bending Frequencies of a Rotating Beam, Hsu Lo, J. E. Goldberg, and J. L. Bogdanoff. (60—APMW-5)

551 Forced Lateral Vibration of a Uniform Cantilever Beam With Internal and External Damping, Ho Chong Lee. (60—APM-15)

557 A Method for Investigating Certain Eigenvalue Problems of the Buckling and Vibration of Plates, H. D. Conway and A. W. Leissa. (60—APM-25)

559 Dynamical Response of an Elastic Half-Space to Tangential Surface Loadings, Chi-Chang Chao. (60—APM-29)

568 On a Special Class of Self-Sustained Oscillations, K. Klotter and E. Kreyszig.

575 Random Excitation of a Loaded Nonlinear String, T. K. Caughey. (60—APMW-1)

579 On the Use of Nonsinusoidal Approximating Functions for Nonlinear Oscillation Problems, K. Klotter and P. R. Cobb.

Tensile and Short-Time Creep Properties of N-155 Alloy Sheet[1]

EUGENE C. BERNETT
Assistant Supervisor, Materials and Process Section, The Marquardt Corporation, Van Nuys, Calif.

The special test equipment and procedures developed for the evaluation of ultrashort-time elevated-temperature mechanical properties of materials are described herein. The tensile and high-stress creep data obtained for Haynes N-155 alloy sheet are reported. At rapid strain rates, the high-temperature tensile strengths were up to 300 per cent higher than those obtained at conventional testing speeds. At high temperatures and high stress levels, however, creep rates were shown to be extremely rapid, and in most cases, the creep strength became a critical factor. The application of rapid-strain-rate tensile data and high-stress creep data in short-life designs is discussed.

IT IS WELL RECOGNIZED that efficient designs of structures which operate at high temperature must be based on material test data which simulate the actual service conditions. Most of the elevated temperature data reported in the literature today are intended for long life, relatively steady-state conditions such as those encountered in steam and gas-turbine power plants. In most missiles and in many "one-mission" type applications, the service life is very short. Therefore materials can be used at much higher temperatures and stresses than those normally considered permissible for prolonged service [1, 2].[2] Where higher stresses can be used, components can of course be made lighter. This is especially desirable in missile applications.

Special test equipment [3] has been developed by the Marquardt Corporation to evaluate the high-temperature properties of materials used in missile ramjet engines and new testing techniques are being used. The universal elevated-temperature machine discussed in [3] can be used for controlled rapid-rate tensile and compression testing (from zero to 0.2 in/in/sec), creep testing, low-frequency fatigue testing, and thermal-cycling tests. It is also possible to program simulated-service trajectories of load and temperature and to make continuous measurements of the resulting strain. Self-resistance heating is used which permits heating rates in excess of 200 F/sec and test temperatures over 3500 F.

A test program to evaluate the tensile and short-time creep properties of N-155 sheet material (20 Cr, 20 Ni, 20 Co, 40 Fe) is reported herein. The test procedures are described in detail and compared with conventional methods. The short-time tensile properties at 75 to 2100 F were determined at various strain rates ranging from 0.0001 to 0.2 in/in/sec. The individual test results are tabulated and typical data are shown graphically. The short-time creep properties were evaluated at temperatures varying from 1600 to 2100 F and the results are presented as standard-creep versus time plots and also in isochronous curve form.

The tensile and the constant-stress data obtained in this program were analyzed and a direct comparison of tensile and creep strength at temperatures of 1600 to 2100 F is shown, Table 11. The relative importance of these two properties is discussed and the conditions at which these data can be used to maximum advantage in short-life, high-temperature designs are pointed out.

Tensile-Testing Procedures

Conventional tensile-testing procedures cannot be used satisfactorily at ultrahigh temperatures. The heating rate, the hold time at temperature, and the strain rate during the test are three extremely important variables which greatly influence the measured tensile properties [4]. The test machine used in obtaining the data which are reported precisely controls these variables. Thus more realistic results are obtained.

The testing speed commonly used in industry for determining tensile strength is such that the yield strength is reached in approximately 30 seconds. The speed is usually increased beyond the yield point so that rupture occurs about a minute or two later. This testing speed corresponds roughly to strain rates of 0.0001 in/in/sec through yield and then 0.002 in/in/sec to rupture. At the high temperatures where the creep process is operative, a large amount of creep occurs during such a slow-rate tensile test and it therefore reduces the tensile strength. Furthermore, the autographic stress-strain record is badly distorted and cannot be accurately interpreted. To minimize the effects of creep deformation during tensile tests, a testing speed of 0.001 in/in/sec through the yield was adopted as a standard for the work described. Beyond the yield, the strain rate is increased to 0.01 in/in/sec. Thus yield is reached in about 3 seconds and rupture occurs about 10 seconds later. The standard speeds adopted for the very-rapid-rate tests were 0.01 in/in/sec through yield and then 0.1 in/in/sec to rupture.

The specimens tested at elevated temperature were heated by electrical self-resistance at a controlled rate of 200 F/sec. Three 36-gage (0.005 in.) chromel-alumel thermocouples were spot welded directly to the test specimen and the temperature of the test portion was maintained within 1 per cent of the set point. The hold time at temperature was 15 min, thus simulating a typical service condition.

The strain measurements were made with a separable 2-in-gage differential transformer extensometer attached to the test portion of the specimen. Spring-loaded extensometer attachment points were provided so that the extensometer would stay on the specimen as it necked down. This permitted recording the full stress-strain curve.

Creep-Testing Procedures

The short-time creep data reported were determined for the

[1] This research was supported in whole or in part by the United States Air Force under Contract No. AF 33(600)-33517, monitored by the Materials Laboratory, Wright Air Development Center, Wright-Patterson Air Force Base, Ohio.

[2] Numbers in brackets designate References at end of paper.

Contributed by the Production Engineering Division and presented at the Annual Meeting, Atlantic City, N. J., November 29–December 4, 1959, of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS.

NOTE: Statements and opinions advanced in papers are to be understood as individual expressions of their authors and not those of the Society. Manuscript received at ASME Headquarters, April 23, 1959. Paper No. 59—A-27.

type of application in which high stresses and temperatures are encountered for short times. The very rapid creep rates obtained at these high stresses and temperature would not permit the use of standard, long-time creep-testing techniques. Therefore the methods to be described were adopted.

The specimens were resistance heated to the test temperature at the same rate that was used in the tensile tests, namely, 200 F/sec, and the thermal gradient was corrected to within 1 per cent. The required stress was applied rapidly, 60 seconds after the test temperature was reached. A rapid loading rate (approximating a strain rate of 0.1 in/in/sec) was used in order to avoid any plastic deformation during the loading period. The plastic creep-strain curve (up to 4 per cent strain or 900 seconds—whichever occurred first) was recorded from the moment that the required stress level was reached. Strain was measured with the extensometer attached directly to the test portion of the specimen, as in the tensile tests. The creep curves were recorded on a Sanborn oscillograph recorder at chart speeds such that the curve could be read at 0.1-sec intervals for the full duration of the test.

In the high-stress creep tests, rupture occurred before the 900-second time limit was reached, and these values were recorded. At low stress levels where rupture did not occur within 900 seconds, the creep test was interrupted, and the residual ultimate strength (at temperature) was determined. The load was increased manually at a speed corresponding to a strain rate of about 0.01 in/in/sec until tensile failure occurred. The ultimate strength values were based on the original cross-sectional area.

Experimental Results

Test Material. The data which are reported were obtained from tests of mill-annealed Haynes N-155 alloy sheet, Aeronautical Material Specification (AMS) 5532B, supplied by the Haynes Stellite Company. All specimens were machined to the dimensions shown in Fig. 1 and they were all longitudinal except where otherwise noted. The certified chemical composition of each heat of material used in these tests is shown in the following tabulation.

Tensile-Test Results. The tensile properties of the subject material were determined over the range from room temperature to 2100 F (1149 C). In addition to the tests at the standard strain rate of 0.001 in/in/sec through yield, checks were made at a strain rate of 0.0001 in/in/sec for comparison with published conventional tensile data. The yield strength at a rapid strain rate of 0.01 in/in/sec was also determined. The ultimate strengths were determined at strain rates of 0.01, 0.1, and 0.2 in/in/sec. In all cases, the heating rate was 200 F/sec and the hold time at temperature was 15 minutes. The individual tensile-test results are listed in Tables 1 through 5. The typical yield-strength, ultimate-strength, elongation, Young's-modulus, and stress-strain curves derived from the tabulated data are shown in Figs. 2, 3, 4, and 5.

Creep-Test Results. The short-time creep properties at various stresses were evaluated at temperatures of 1600, 1800, 1900, 2000, and 2100 F (871, 982, 1038, 1093, and 1149 C). Standard flat specimens, as shown in Fig. 1, were used. The highest stress levels which were used resulted in creep rates of approximately 0.5 per cent/sec (1800 per cent/hr). Several tests (as required) were made at each stress level and temperature in order to establish the creep scatter band. The results which were obtained are tabulated in Tables 6 through 10. The typical loading-strain curves are shown in Fig. 6. Plots of maximum plastic creep strain versus time are shown in Figs. 7 to 11 and the same data are shown in isochronous-creep-curve form in Figs. 12 to 16.

Composite Results. The typical tensile and creep strengths at 1600 to 2100 F (871 to 1149 C) are compared in Table 11. The tensile values listed in Table 11 were taken from the curves shown in Figs. 2 and 3, the creep data were taken from Figs. 12 to 16, and the rupture data were taken from Table 6 to 10.

Discussion of Results

Tensile Properties. The curves shown in Figs. 2, 3, 4, and 5 indicate the typical tensile properties of annealed N-155 sheet material, based on the individual test values listed in Tables 1 through 5. The scatter in the measured properties from heat to heat is small, which indicates that a consistent quality level is being maintained in the production process. The typical-tensile-property curves are therefore considered representative and they are suitable for design purposes. Appropriate safety factors (based on the specific application) should be used in all cases.

Element	AMS 5532B Requirements	Heat Number						
		M953	M1492	M1675	M5015	M5115	M5148	M5171
C	0.08 to 0.16	0.12	0.09	0.16	0.12	0.10	0.11	0.10
Mn	1.0 to 2.0	1.73	1.59	1.48	1.68	1.64	1.38	1.70
Si	1.0 max.	0.60	0.55	0.55	0.62	0.69	0.60	0.73
P	0.04 max.	0.020	0.014	0.012	0.012	0.014	0.017	0.015
S	0.03 max.	0.013	0.017	0.020	0.016	0.012	0.011	0.014
Cr	20 to 22.5	21.0	20.9	22.1	21.7	22.1	21.3	21.6
Ni	19 to 21	19.9	20.1	20.1	19.9	20.0	20.0	20.1
Co	19.5 to 21	19.7	19.9	19.8	19.7	19.4	20.1	20.0
Mo	2.5 to 3.5	3.08	3.04	2.91	3.08	3.11	2.96	3.07
W	2.0 to 3.0	2.62	2.53	2.56	2.46	2.40	2.44	2.42
Cb+Ta	0.75 to 1.25	0.86	0.95	0.80	1.14	0.95	0.88	1.01
N	0.1 to 0.2	0.12	0.12	0.10	0.12	0.12	0.12	0.13
Fe	Balance	Bal.	Bal.	Bal.	Bal.	Bal.	Bal.	Bal.

Table 1 Tensile properties of N-155 alloy sheet

Strain Rate: 0.0001 in./in./sec to YS then 0.01 in./in./sec to Rupture*

Specimen No.	Heat No.	Gage (in.)	Temp. (°F)	Soak (min)	Prop. Limit (Ksi)	0.2% YS (Ksi)	UTS (Ksi)	Elong. (% in 2 in.)	E ($\times 10^6$ psi)
M587	M5171	0.063	RT	--	39.0	61.7	117.0	41	30.0
M604	M5171	0.063	RT	--	41.1	57.6	114.4	46	32.4
M605	M5171	0.063	1000	15	23.8	37.6	88.8	42	21.0
M606	M5171	0.063	1000	15	25.1	37.7	88.0	42	22.0
M607	M5171	0.063	1300	15	26.6	32.9	--	35	14.0
M608	M5171	0.063	1300	15	23.4	33.0	73.0	30	15.5
M617	M5171	0.063	1300	15	23.8	33.0	71.3	29	20.2
Q137	M5148	0.063	1300	15	26.4	31.0	74.0	48	20.6
Q138	M5148	0.063	1300	15	--	30.7	73.8	42	17.0
M609	M5171	0.063	1600	15	24.9	27.4	45.4	22	21.0
M610	M5171	0.063	1600	15	25.0	27.3	44.1	22	23.0
Q139	M5148	0.063	1600	15	20.1	24.8	45.0	28	16.8
Q140	M5148	0.063	1600	15	18.8	26.2	44.4	25	15.7
M611	M5171	0.063	1800	15	9.2	12.3	24.2	23	7.9
M612	M5171	0.063	1800	15	8.5	10.8	23.0	28	13.4
M618	M5171	0.063	1800	15	8.8	11.7	24.4	21	14.4
Q141	M5148	0.063	1800	15	12.3	14.2	29.6	22	13.3
Q142	M5148	0.063	1800	15		13.9	25.8	23	16.6
M613	M5171	0.063	2000	15	3.0	4.5	11.8	23	9.0
M614	M5171	0.063	2000	15	3.0	4.9	12.4	23	9.3
Q143	M5148	0.063	2000	15	--	5.5	16.3	23	7.7
Q144	M5148	0.063	2000	15	3.5	5.7	16.7	20	10.1
M615	M5171	0.063	2100	15	2.2	3.1	9.0	25	6.2
M616	M5171	0.063	2100	15	1.3	3.4	9.6	25	5.9
Q145	M5148	0.063	2100	15	3.3	4.1	14.7	20	5.9
Q146	M5148	0.063	2100	15	3.8	4.3	13.6	24	6.6

* Note: The 0.0001 in./in./sec strain rate corresponds to the speed used in conventional tensile tests.

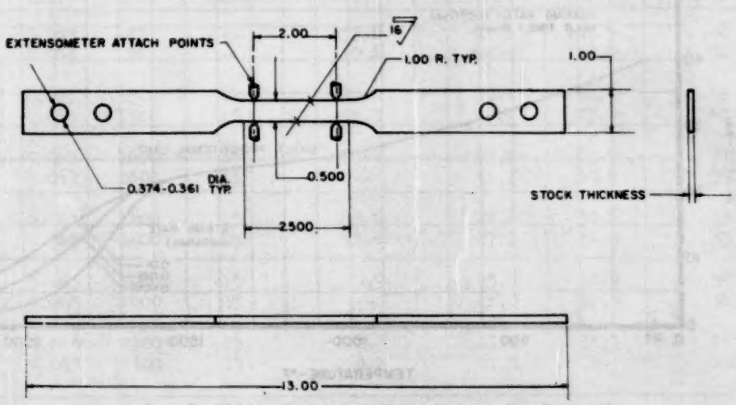


Fig. 1 Standard flat tensile and creep-test specimen

Table 2 Tensile properties of N-155 alloy sheet

Strain Rate: 0.001 in./in./sec to YS then 0.01 in./in./sec to Rupture

Specimen No.	Heat No.	Gage (in.)	Temp. (°F)	Soak (min)	Prop. Limit (Ksi)	0.2% YS (Ksi)	UTS (Ksi)	Elong. (% in 2 in.)	E ($\times 10^6$ psi)
M591	M5171	0.063	RT	--		61.5	113.4	46	33.2
Q2	M5148	0.063	RT	--	40.7	64.0	117.0	49	28.4
Q17	M5148	0.063	RT	--	22.1	59.8	117.9	46	--
Q18*	M5148	0.063	RT	--	26.5	57.2	115.4	46	30.0
M598	M5171	0.063	1000	15	25.8	35.0	87.2	38	23.6
M599	M5171	0.063	1000	15	25.0	32.0	89.1	44	33.8
Q3	M5148	0.063	1000	15	27.9	33.4	87.8	45	20.6
Q5	M5148	0.063	1000	15	25.4	31.9	86.0	49	21.7
M600	M5171	0.063	1300	15	24.3	32.2	--	39	16.2
M601	M5171	0.063	1300	15	--	--	73.6	31	--
Q6	M5148	0.063	1300	15	23.6	31.6	75.8	42	20.9
Q9	M5148	0.063	1300	15	20.5	31.1	76.4	42	20.5
M602	M5171	0.063	1600	15	21.5	30.1	41.4	22	14.8
M603	M5171	0.063	1600	15	24.1	31.6	42.4	22	18.0
Q7	M5148	0.063	1600	15	18.7	28.5	--	34	12.6
Q8	M5148	0.063	1600	15	22.0	29.0	--	35	14.7
M592	M5171	0.063	1800	15	13.1	15.9	25.6	23	14.9
M593	M5171	0.063	1800	15	16.1	16.8	25.2	24	13.9
Q10	M5148	0.063	1800	15	13.4	17.6	26.9	26	11.7
Q15	M5148	0.063	1800	15	17.7	18.6	28.2	24	13.3
M594	M5171	0.063	2000	15	7.0	8.4	13.2	25	6.1
M595	M5171	0.063	2000	15	5.9	8.1	13.7	27	7.1
Q11	M5148	0.063	2000	15	8.2	9.5	15.1	20	8.6
Q12	M5148	0.063	2000	15	5.6	7.6	13.1	25	8.7
Q16	M5148	0.063	2000	15	8.5	8.6	15.3	27	9.0
M596	M5171	0.063	2100	15	5.6	6.4	11.1	20	3.8
M597	M5171	0.063	2100	15	5.4	6.5	11.5	24	4.3
Q13	M5148	0.063	2100	15	5.1	6.6	11.9	37	6.0
Q14	M5148	0.063	2100	15	4.1	6.4	11.3	23	6.2

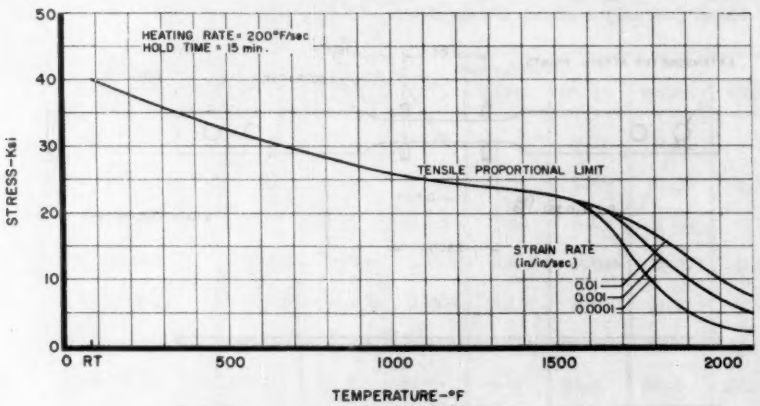


Fig. 2 Tensile-proportional limit curves for N-155 alloy sheet

Table 3 Tensile properties of N-155 alloy sheet

Strain Rate: 0.001 in/in/sec to YS then 0.2 in/in/sec to Rupture

Specimen No.	Heat No.	Gage (in.)	Temp. (°F)	Soak (min)	Prop. Limit (Ksi)	0.2% YS (Ksi)	UTS (Ksi)	Elong. (% in 2 in.)	E ($\times 10^6$ psi)
M239	M1492	0.078	RT	--	42.2	60.2	111.6	48	30.0
M619	M5171	0.063	RT	--	33.1	60.0	116.8	44	30.6
M242	M1492	0.078	1300	15	21.8	33.0	74.5	40	18.8
M621	M5171	0.063	1300	15	23.0	32.5	76.2	40	20.0
M241	M1492	0.078	1800	15	14.5	19.1	36.7	27	13.7
M622	M5171	0.063	1800	15	15.5	16.8	33.2	21	14.0
M243	M1492	0.078	2100	15	3.9	6.1	15.8	30	6.6

Table 4 Tensile properties of N-155 alloy sheet

Strain Rate: 0.01 in./in./sec from Start to Rupture

Specimen No.	Heat No.	Gage (in.)	Temp. (°F)	Soak (min)	Prop. Limit (Ksi)	0.2% YS (Ksi)	UTS (Ksi)	Elong. (% in 2 in.)	E ($\times 10^6$ psi)
M245	M1492	0.078	RT	--	41.5	63.2	114.2	42	32.1
M244	M1492	0.078	1300	15	27.8	32.0	72.5	36	16.6
M240	M1492	0.078	1600	15	20.5	30.6	--	26	16.4
M246	M1492	0.078	1800	15	14.5	27.3	27.3	18	--
M620	M5171	0.063	2100	15	6.9	10.5	10.5	22	--

Table 5 Tensile properties of N-155 alloy sheet

Strain Rate: 0.01 in./in./sec to YS then 0.1 in./in./sec to Rupture

Specimen No.	Heat No.	Gage (in.)	Temp. (°F)	Soak (min)	Prop. Limit (Ksi)	0.2% YS (Ksi)	UTS (Ksi)	Elong. (% in 2 in.)	E ($\times 10^6$ psi)
Q125	M5148	0.063	RT	--	33.8	62.5	117.9	48	33.0
Q126	M5148	0.063	RT	--	39.6	62.9	118.9	52	30.1
Q127	M5148	0.063	1300	15	25.0	30.4	76.1	42	
Q128	M5148	0.063	1300	15	26.3	29.9	75.4	42	20.7
Q129	M5148	0.063	1600	15	23.4	28.6	53.5	34	19.6
Q130	M5148	0.063	1600	15	--	31.4	50.9	34	--
Q131	M5148	0.063	1800	15	15.0	27.8	35.6	33	13.8
Q132	M5148	0.063	1800	15	14.8	27.1	34.9	33	11.2
Q133	M5148	0.063	2000	15	9.9	15.5	20.4	38	8.3
Q134	M5148	0.063	2000	15	--	18.0	22.8	38	9.3
Q135	M5148	0.063	2100	15	8.3	11.6	15.7	27	6.3
Q136	M5148	0.063	2100	15	8.2	11.5	--	20	5.7

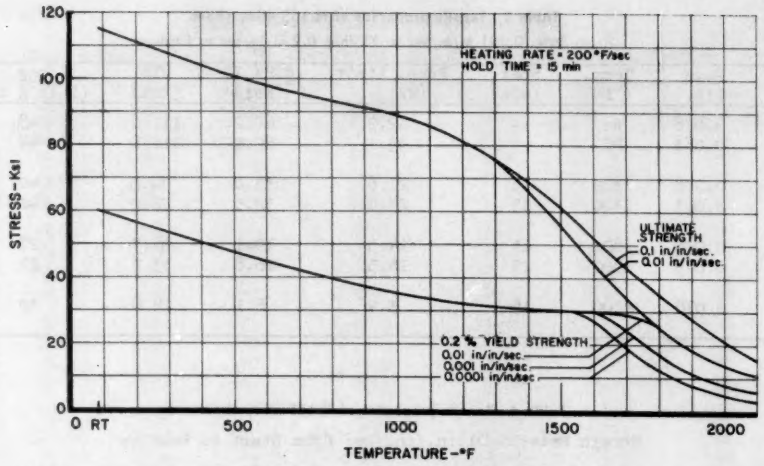


Fig. 3 Tensile-strength curves for N-155 alloy sheet

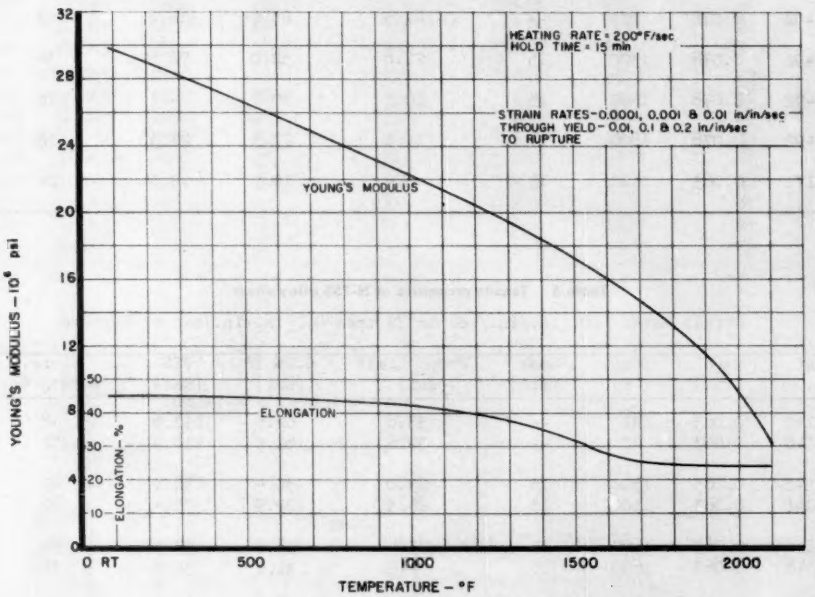


Fig. 4 Tensile elongation and modulus curves for N-155 alloy sheet

The effect of testing speed on the elevated-temperature tensile-proportional limit stress, 0.2 per cent yield strength, and ultimate strengths of the subject material are readily apparent from the curves shown in Figs. 2, 3, and 5. At strain rates between 0.0001 and 0.01 in/in/sec, the proportional limit stress and 0.2 per cent yield strengths are both strain-rate sensitive above 1500 F (816 C). At testing speeds between 0.01 and 0.2 in/in/sec, the ultimate strength is strain-rate sensitive above 1200 F (649 C). No consistent differences were noted in the elongation and modulus values obtained at the different testing speeds.

It might be noted that, in the tensile tests reported here, the strain rate was normally increased (by a factor of 10 or 100) once the yield strength was exceeded. At elevated temperatures, where the yield was strain-rate sensitive, the instantaneous increase in testing speed resulted in a second modulus line on the autographic stress-strain record. This second straight line on the stress-strain diagram provided an accurate check on the Young's-modulus value determined from the initial portion of the curve.

The hold time at temperature in the tensile tests reported here was 15 minutes. In a subsequent series of tests, hold times of 2 and 30 minutes were investigated. The tensile properties obtained were essentially the same as those reported here. The typical 15-min-hold-time tensile-property curves shown in Figs. 2, 3, 4, and 5 are therefore considered applicable for hold times ranging from zero to at least 30 minutes.

The increased-tensile-strength values obtained in the fast-rate tests (0.01 in/in/sec through yield then 0.1 in/in/sec to rupture) should not be indiscriminately substituted for the currently employed conventional allowable strength. The higher-tensile-strength values should be used only in those cases in which the service stresses are applied at comparable rates and the peak

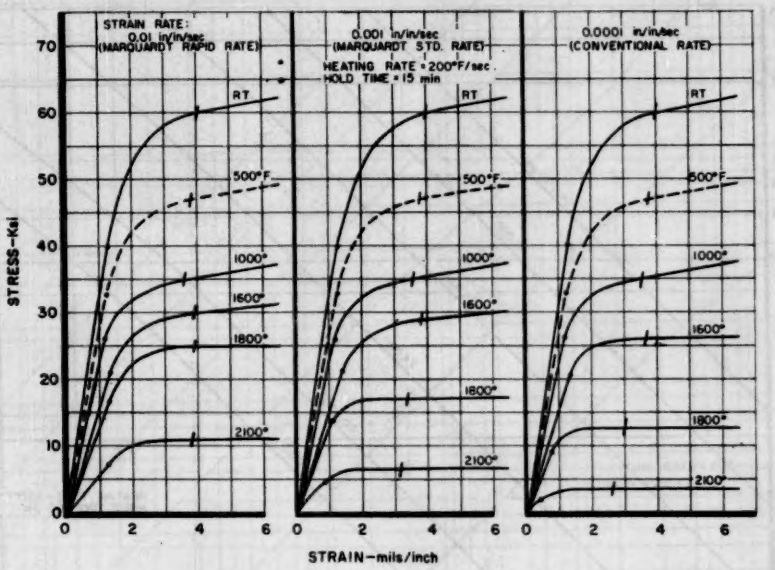


Fig. 5 Tensile stress-strain curves for N-155 alloy sheet

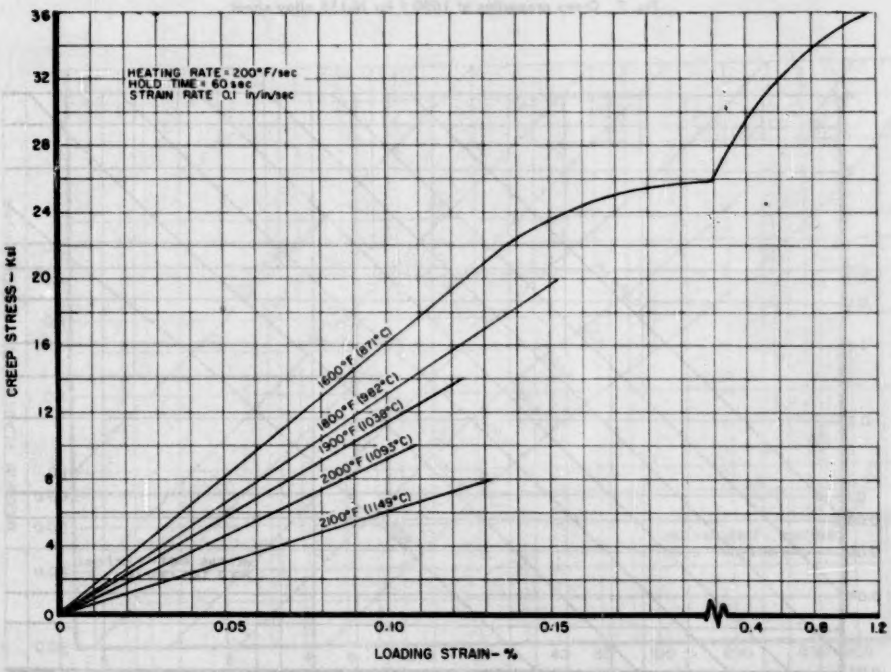


Fig. 6 Creep-test loading-strain plots for N-155 alloy sheet

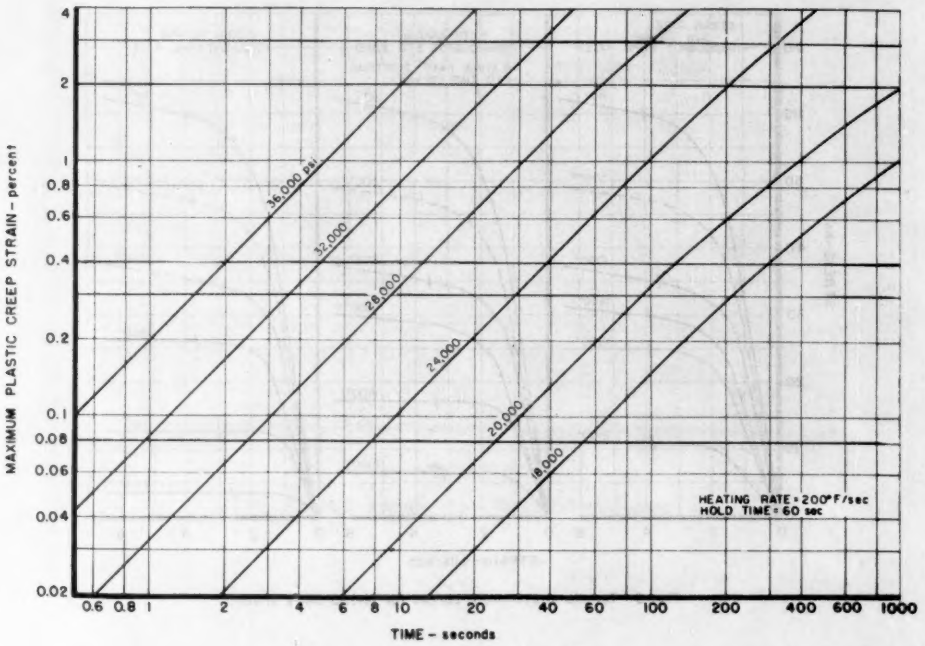


Fig. 7 Creep properties at 1600 F for N-155 alloy sheet

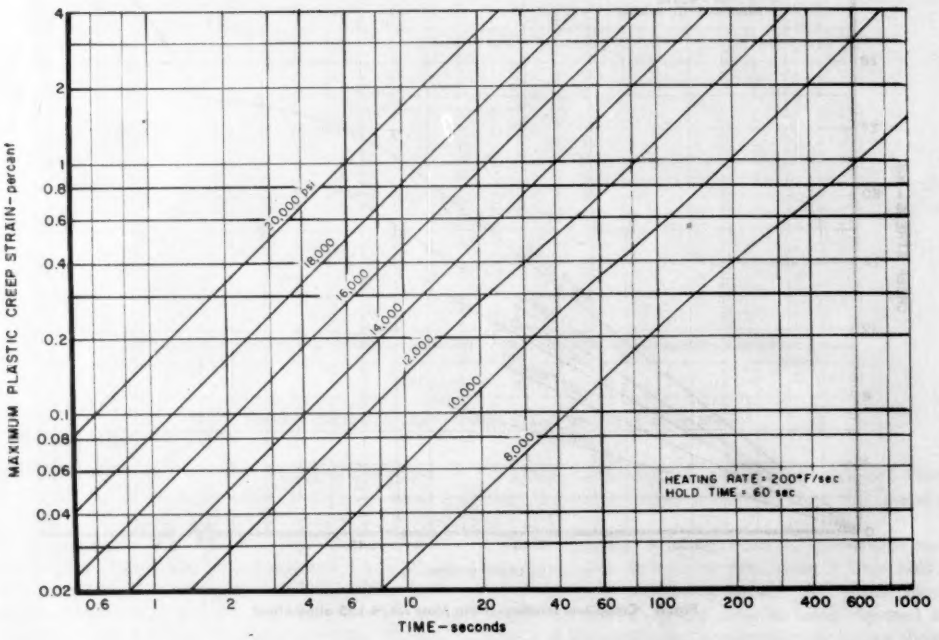


Fig. 8 Creep properties at 1800 F for N-155 alloy sheet

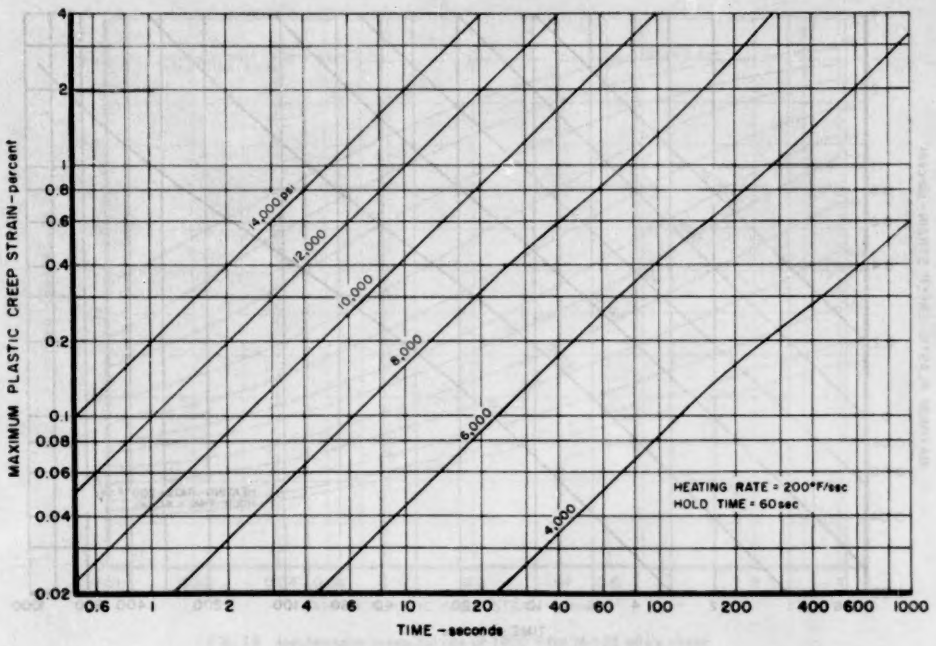


Fig. 9 Creep properties at 1900 F for N-155 alloy sheet

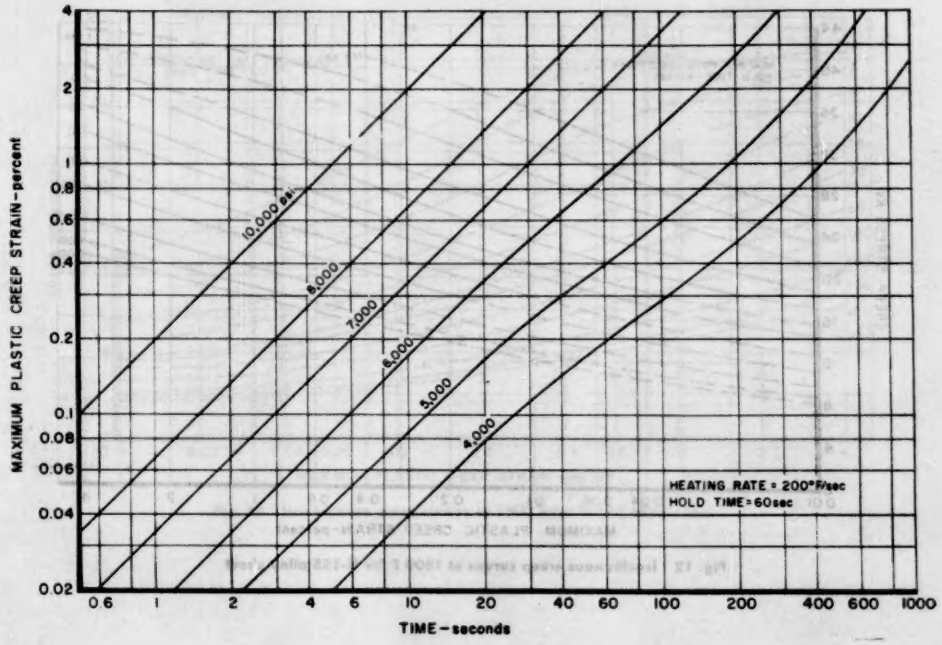


Fig. 10 Creep properties at 2000 F for N-155 alloy sheet

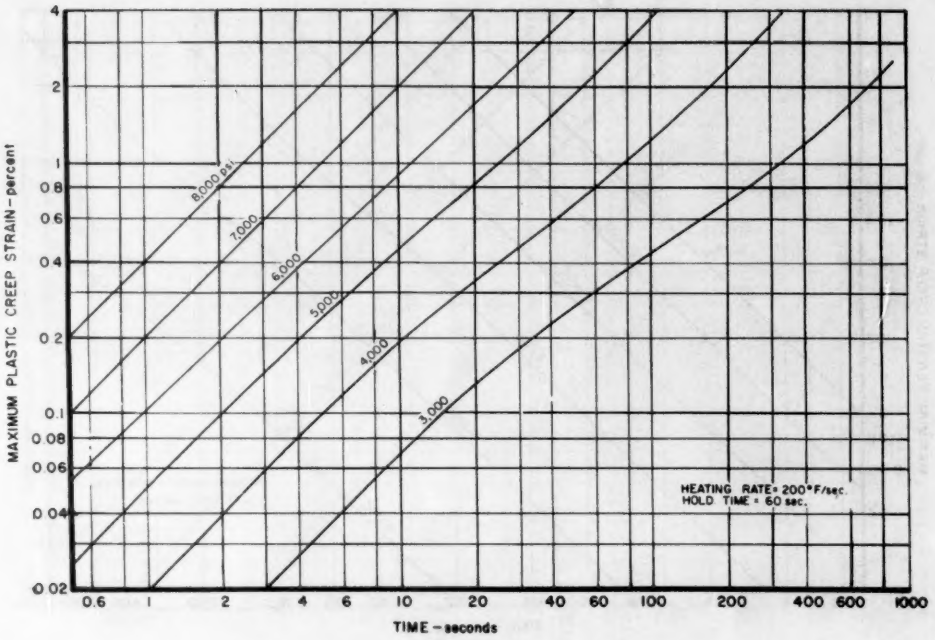


Fig. 11 Creep properties at 2100 F for N-155 alloy sheet

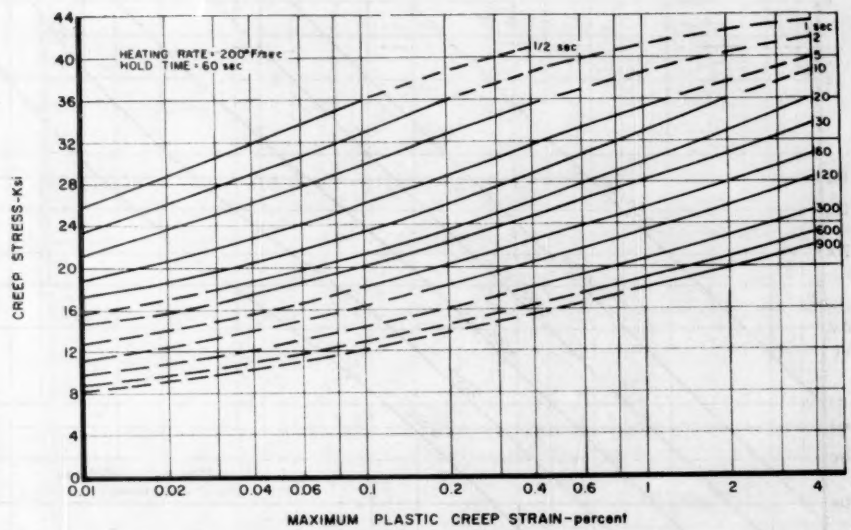


Fig. 12 Isochronous creep curves at 1600 F for N-155 alloy sheet

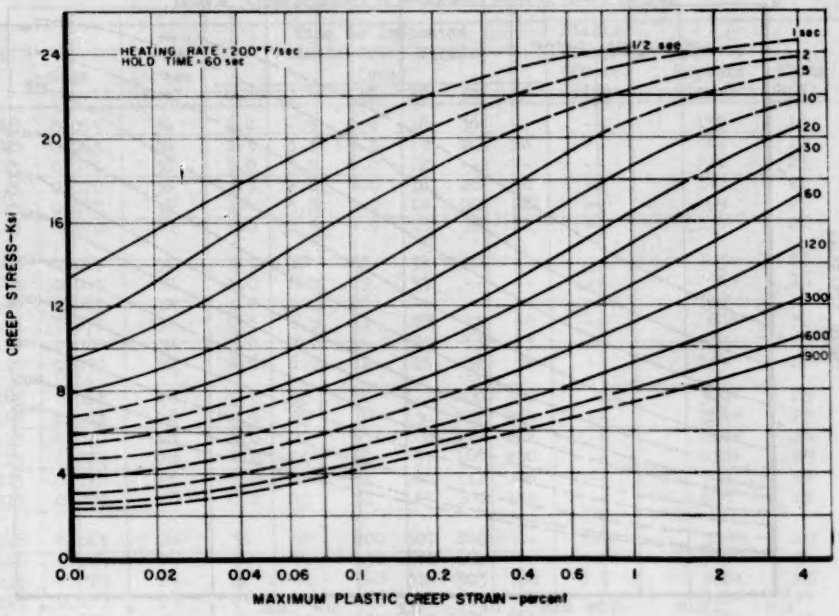


Fig. 13 Isochronous creep curves at 1800 F for N-155 alloy sheet

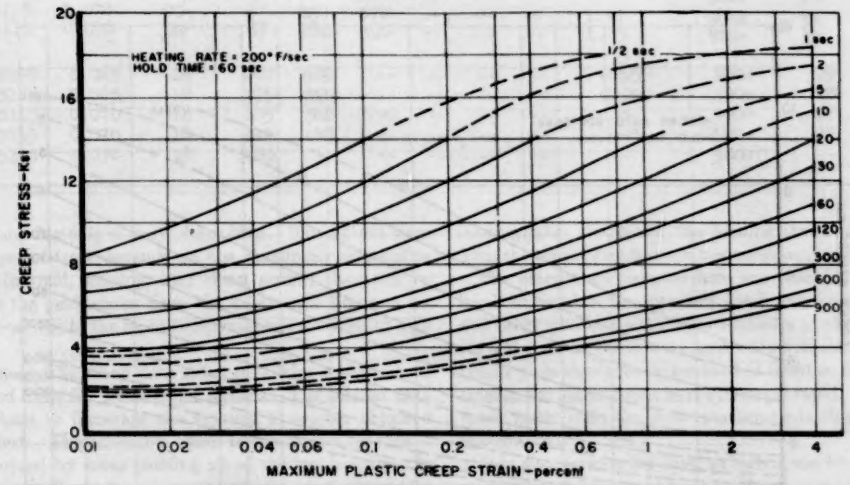


Fig. 14 Isochronous creep curves at 1900 F for N-155 alloy sheet

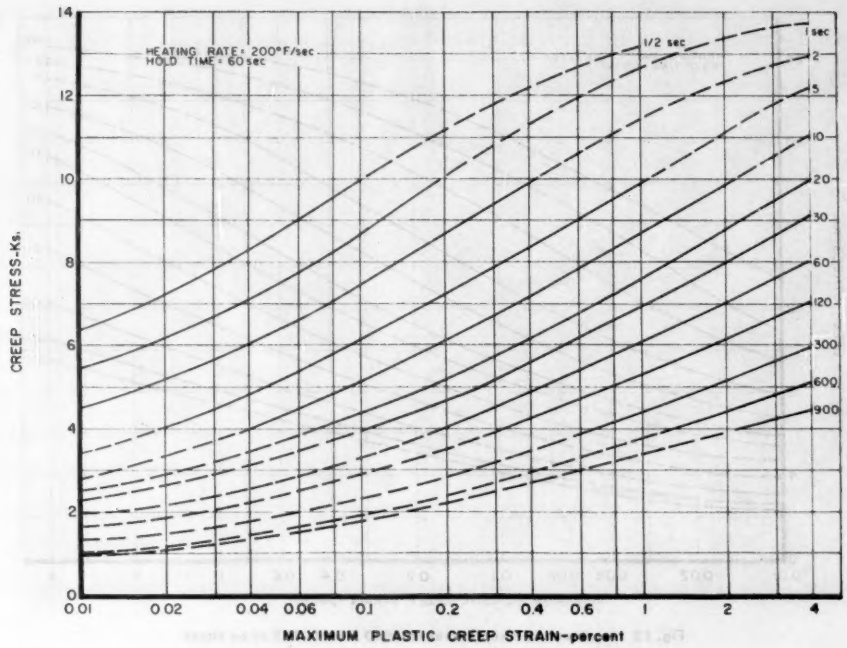


Fig. 15 Isochronous creep curves at 2000 F for N-155 alloy sheet

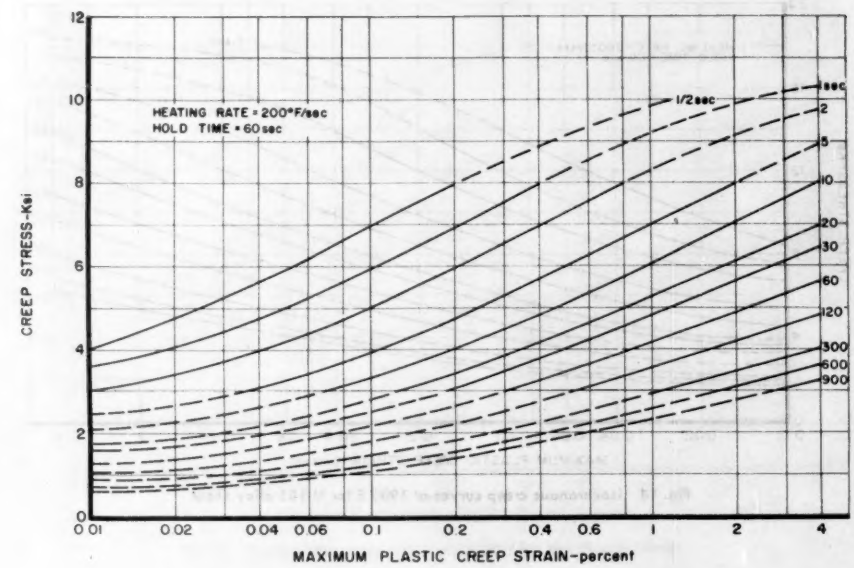


Fig. 16 Isochronous creep curves at 2100 F for N-155 alloy sheet

Table 6 Creep properties of N-155 alloy sheet at 1600 F (871 C)

Specimen No.	Heat No.	Gage (in.)	Creep Stress (Ksi)	Time to Indicated Plastic Creep Strain (sec)						Plastic Creep in 900 sec (%)	Time to Rupture (sec)	Elong. (%)	Residual UTS at 1600°F (Ksi)
				0.2%	0.5%	1%	2%	3%	4%				
Q183	M5148	0.063	36	1.8	4.5	9.0	18	36	--	--	159	18	--
Q184	M5148	0.063	36	1.3	3.3	6.5	13	20	26	--	82	18	--
R789	M5115	0.070	36	1.0	3.0	6.0	12	--	--	--	116	25	--
R812	M5115	0.070	36	1.5	4.0	8.0	16	28	38	--	173	25	--
R830	M5115	0.070	36	1.4	3.2	7.0	14	20	28	--	99	25	--
R831	M5115	0.070	36	1.0	2.3	4.9	10	16	22	--	116	27	--
Q180	M5148	0.063	32	3.5	10	23	54	90	--	--	103+	27	56.0
Q181	M5148	0.063	32	3.0	8.5	19	45	--	--	--	96+	18	--
Q185	M5148	0.063	32	3.0	9.0	20	--	--	--	--	339	18	--
R787	M5115	0.070	32	2.5	6.5	13	28	43	60	--	60+	32	49.6
R806	M5115	0.070	32	3.0	7.8	16	33	53	73	--	334	25	--
R819	M5115	0.070	32	2.6	6.2	12	24	38	52	--	244	26	--
Q178	M5148	0.063	28	9.5	26	56	112	200	300	--	300+	25	50.2
Q179	M5148	0.063	28	9.0	25	76	136	207	284	--	284+	25	53.2
R785	M5115	0.070	28	8.0	22	45	100	159	222	--	222+	26	52.5
R786	M5115	0.070	28	8.0	20	40	97	166	230	--	230+	25	50.0
R790	M5115	0.070	28	6.0	15	31	65	111	156	--	759	27	--
R805	M5115	0.070	28	6.2	15	31	67	109	144	--	625	29	--
Q175	M5148	0.063	24	36	90	200	507	840	--	3.15	900+	20	51.9
Q176	M5148	0.063	24	28	74	130	380	682	--	3.62	900+	24	56.0
R771	M5115	0.070	24	18	45	140	316	503	700	5.0	900+	20	50.6
R782	M5115	0.070	24	24	71	168	410	695	--	3.8	900+	27	52.8
R783	M5115	0.070	24	24	59	139	313	486	665	--	665+	25	52.0
Q172	M5148	0.063	20	110	320	675	--	--	--	1.27	900+	22	52.5
Q173	M5148	0.063	20	169	408	780	--	--	--	1.17	900+	30	54.5
R766	M5115	0.070	20	90	206	430	840	--	--	2.35	900+	24	49.5
R768	M5115	0.070	20	63	150	300	600	--	--	3.0	900+	28	53.1
R780	M5115	0.070	20	67	220	570	--	--	--	1.47	900+	32	47.0
R781	M5115	0.070	20	87	280	595	--	--	--	1.35	900+	30	52.7
Q167	M5148	0.063	18	240	680	--	--	--	--	0.65	900+	29	57.0
Q177	M5148	0.063	18	175	390	--	--	--	--	0.90	900+	20	--
R765	M5115	0.070	18	105	360	790	--	--	--	1.5	900+	27	50.0
R778	M5115	0.070	18	144	440	--	--	--	--	0.95	900+	27	54.0
R779	M5115	0.070	18	180	--	--	--	--	--	--	900+	--	53.0

stress is retained for only a very short time. The actual time delay at the peak stress depends on the maximum permissible deformation (thermal, loading, and creep strain) that can be tolerated. As the peak stress rises, the delay time becomes increasingly critical due to the rapid creep rates encountered at high stress levels.

Creep Properties. The creep data listed in Tables 6 through 10 were condensed from the oscillograph records. The tabular data are included here to illustrate the amount of scatter obtained from test to test. The allowable creep curves shown in Figs. 7 to 16 were derived by cross plotting all of the test results. In every case, they indicate the maximum plastic creep strain at a given time, stress, and temperature in question. It should be realized, however, that these curves are based on test results from a limited number of heats of material and they therefore are not "guaranteed" properties. Nevertheless, the creep strength allowables shown in Figs. 7 to 16 are conservative values and minimum safety factors therefore should suffice in all but the most critical design applications.

The rate of load application in the creep tests reported here was very rapid, corresponding to a strain rate of approximately 0.1 in/in/sec. Except in the 1600-F creep tests at stresses above 20,000 psi, the high-strain-rate tensile-proportional limit stress of this material at each creep-test temperature is above the maximum creep stresses used. Therefore the loading strains were small and entirely elastic, Fig. 6. In design cases in which the loading rates are slower, the loading strain may be determined from the tensile stress-strain curves shown in Fig. 5.

The creep tests reported here were carried out at high stress levels to simulate the transient conditions encountered in missiles and other similar short-life, one-mission applications. The highest test stresses which were used were well above the 0.2 per cent tensile yield strengths determined at 0.001 in/in/sec. The creep rates at the high stresses were extremely rapid. In fact, creep rates much faster than the slow-rate tensile-testing strain rates were encountered.

It is also noteworthy that, at high stress levels, the creep rates up to at least 4 per cent plastic deformation are constant; that is, the creep curves are straight lines. These high-stress creep curves can therefore be readily extrapolated to very short-time intervals. Such short-time creep data would be particularly useful for critical applications in which high transient stresses are encountered.

The post-creep residual tensile strengths determined here compare favorably with the typical ultimate strengths determined for as-received annealed material, Fig. 3. It should be noted that the post-creep tensile strengths were determined by increasing the load manually. The strain rate was therefore not closely controlled and this factor, plus the damage due to creep, would largely account for the scatter which was obtained in the measured residual strengths.

In the graphical presentation of the creep data, the loading strain and the plastic-creep strain curves are shown separately.

Table 7 Creep properties of N-155 alloy sheet at 1800 F (982 C)

Specimen No.	Heat No.	Gage (in.)	Creep Stress (Ksi)	Time to Indicated Plastic Creep Strain (sec)						Plastic Creep in 900 sec (%)	Time to Rupture (sec)	Elong. (%)	Residual UTS at 1800°F (Ksi)
				0.2%	0.5%	1%	2%	3%	4%				
Q119	M5148	0.063	20	1.6	3.4	6.0	13	19	25	--	75	22	--
Q120	M5148	0.063	20	2.0	5.0	10	18	25	32	--	87	22	--
Q122	M5148	0.063	20	1.2	3.1	6	12	17	23	--	70	24	--
Q91	M5148	0.063	18	5.0	10	18	36	45	63	--	200+	--	--
Q92	M5148	0.063	18	2.8	6.4	12	24	35	46	--	100+	18	--
Q93	M5148	0.063	18	6.0	12	21	37	52	65	--	161	19	--
Q94	M5148	0.063	18	3.2	8.0	14	29	42	57	--	141	22	--
Q95	M5148	0.063	16	6.0	13	23	40	60	83	--	241	19	--
Q96	M5148	0.063	16	6.0	13	23	45	67	91	--	226	19	--
Q97	M5148	0.063	16	6.0	13	25	53	80	112	--	425	25	--
Q98	M5148	0.063	16	6.0	12	26	45	67	92	--	328	23	--
Q100	M5148	0.063	14	8.0	24	48	87	125	160	--	500	24	--
Q101	M5148	0.063	14	9.0	20	42	77	117	158	--	563	24	--
Q123	M5148	0.063	12	12	29	64	128	192	263	--	523	15	--
Q124	M5148	0.063	12	14	34	--	161	242	317	--	903	21	--
Q104	M5148	0.063	12	19	49	119	270	428	579	--	900+	19	12.3
Q116	M5148	0.063	12	10	32	69	149	227	--	--	856	20	--
Q107	M5148	0.063	10	23	87	222	516	777	--	3.5	900+	30	29.4
Q108	M5148	0.063	10	32	117	250	435	666	885	--	900+	22	27.8
Q117	M5148	0.063	10	24	78	155	345	558	775	--	900+	28	29.7
Q110	M5148	0.063	10	27	68	142	--	538	850	--	900+	27	28.6
Q118	M5148	0.063	10	55	160	295	510	760	--	3.7	900+	25	27.4
Q111	M5148	0.063	8	150	545	--	--	--	--	0.75	900+	31	24.6
Q112	M5148	0.063	8	73	272	580	--	--	--	1.6	900+	24	24.8
Q113	M5148	0.063	8	140	393	870	--	--	--	1.2	900+	--	26.7
Q114	M5148	0.063	8	83	260	540	--	--	--	--	900+	26	24.3

This method of presentation was used in order to facilitate design calculations of plastic deformation in components subjected to fluctuating stresses. Double logarithmic plots were used for the plastic creep curves and semilogarithmic plots were used for the isochronous curves in order to show the low-strain, short-time values more clearly and to permit easier extrapolation of the data.

Comparative Creep and Tensile Properties. A summarized tabulation of the tensile and short-time creep strengths of annealed N-155 alloy sheet at temperatures of 1600, 1800, 1900, 2000, and 2100 F is shown in Table 11. The tensile-proportional limit stresses, 0.2 per cent yield strength, and ultimate tensile strength determined at slow and rapid rates are listed. The creep behavior and 60 and 600-sec rupture lives at stress levels comparable to those obtained in the tensile tests are also shown. It can be readily observed from the table that creep effects are of major importance at high temperatures even for very short-time service. For example, neglecting safety factors, the allowable working stress at 2100 F, based on the 0.2 per cent yield strength of the subject material, ranges from 3000 to 11,000 psi depending on the strain rate. From a creep standpoint, a sustained stress of 11,000 psi at 2100 F is obviously out of the question. At this stress and temperature, the creep rate would be so rapid that 4 per cent plastic deformation would take place in less than 1 second, and rupture would occur in about 5 seconds. For steady-state conditions, therefore, the working stress should be reduced so that the amount of creep that occurs during the service life does not exceed the permissible limit.

It might be noted that the shape of the isochronous-creep-stress curves, Figs. 12 to 16, provides further evidence that creep at high stress levels is particularly critical. These curves become very flat at high stress levels. Thus a slight increase in stress results in a large increase in creep rate. In the case of highly stressed components which operate at high temperatures, the service time at the peak stress therefore is of utmost importance.

The two creep rates noted in Table 11 (0.001 and 0.0001 in/in/sec) are the same as the two lower strain rates used in the tensile tests reported here. It is noteworthy that the stress required to give the two noted creep rates is roughly equivalent to the 0.2 per cent yield strengths determined at the same strain rates.

As indicated previously, the maximum stress levels evaluated in the creep tests carried out here were such as to produce plastic-creep rates approaching 0.5 per cent/sec. At the lowest creep-test temperature used here (1600 F), stress levels up to 36,000 psi were required. This of course is well beyond the tensile-proportional limit strength and consequently a great deal of plastic deformation (in addition to the elastic) was obtained on loading (Fig. 6). In many cases, the loading deformation would be a critical factor from the design standpoint. Below 1600 F, even higher loading strains would be obtained (at stress levels required to produce creep rates of 0.5 per cent/sec) and the tensile stress-strain behavior becomes increasingly critical. The foregoing indicates that, in applying N-155 alloy to the type of short-life designs under consideration here, 1500 F is approximately the transition temperature below which tensile strength is a governing factor and above which creep strength is critical.

Extensive testing of other aircraft structural materials has shown elevated-temperature behavior similar to that observed for the N-155 alloy. For the most efficient design of elevated-temperature structural components, it is very important to establish and to consider both the tensile and the creep properties of the materials to be used.

Conclusions

1 The tensile properties of annealed N-155 alloy sheet are

Table 8 Creep properties of N-155 alloy sheet at 1900 F (1038 C)

Specimen No.	Heat No.	Gage (in.)	Creep Stress (Ksi)	Time to Indicated Plastic Creep Strain (sec)						Plastic Creep in 900 sec (%)	Time to Rupture (sec)	Elong. (%)	Residual UTS at 1900°F (Ksi)
				0.2%	0.5%	1%	2%	3%	4%				
Q55	M5148	0.063	14	2.0	4.7	9.0	17	23	30	--	86	20	--
Q56	M5148	0.063	14	3.3	7.0	13	25	35	44	--	77	20	--
Q75	M5148	0.063	14	2.5	6.0	10	20	30	39	--	100	21	--
Q90	M5148	0.063	14	2.0	4.0	8.0	15	21	28	--	67	20	--
Q105	M5148	0.063	14	1.5	3.8	7.4	15	22	29	--	81	23	--
M625	M5171	0.063	12	3.0	7.0	14	30	42	54	--	174	--	--
M645	M5171	0.063	12	5.0	17	33	60	--	--	--	336+	--	--
Q58	M5148	0.063	12	5.0	11	22	38	43	68	--	195+	--	--
Q76	M5148	0.063	12	6.0	15	30	57	75	98	--	203	17	--
Q77	M5148	0.063	12	5.0	12	25	42	59	77	--	164	16	--
Q88	M5148	0.063	12	2.5	5.8	10	24	40	56	--	168	17	--
M649	M5171	0.063	10	7.4	19	33	60	88	--	--	108+	--	--
Q62	M5148	0.063	10	4.0	10	23	48	74	98	--	150+	--	--
Q63	M5148	0.063	10	6.0	16	36	60	95	125	--	255+	--	--
M74	M953	0.063	8	12	40	118	--	--	--	--	900+	25	22.4
M75	M953	0.063	8	29	92	180	--	--	--	--	840+	30	30.9
M60	M1675	0.063	8	16	43	78	144	--	--	--	540+	25	--
Q64	M5148	0.063	8	17	46	93	168	250	330	--	782	14	--
Q65	M5148	0.063	8	25	78	165	302	465	619	--	900+	31	25.6
Q66	M5148	0.063	8	29	89	186	322	455	571	--	900+	15	30.1
Q79	M5148	0.063	8	14	41	82	164	250	340	--	793	15	--
G2	M5148	0.063	8	30	66	105	240	--	--	--	--	--	--
H235	M5148	0.063	8	23	55	110	216	--	--	--	--	--	--
H239	M5148	0.063	8	23	54	105	215	--	--	--	--	--	--
M76	M953	0.063	6	68	210	410	810	--	--	--	810+	10	26.8
M77	M953	0.063	6	53	140	345	696	--	--	--	696+	19	26.4
M78	M953	0.063	6	60	180	420	708	--	--	--	708+	16	24.5
Q67	M5148	0.063	6	56	215	446	900	--	--	2	900+	27	19.5
Q68	M5148	0.063	6	60	193	371	757	--	--	2.7	900+	21	19.8
Q69	M5148	0.063	6	58	197	439	--	--	--	1.7	900+	23	24.8
G8	M5148	0.063	6	78	204	420	852	--	--	--	--	--	--
G9	M5148	0.063	6	90	237	417	660	--	--	--	--	--	--
H238	M5148	0.063	6	78	240	525	--	--	--	--	--	--	--
M79	M953	0.063	4	200	610	--	--	--	--	0.75	900+	20	24.6
M80	M953	0.063	4	345	680	--	--	--	--	0.72	900+	20	27.3
M81	M953	0.063	4	230	600	--	--	--	--	0.78	900+	20	22.6
M83	M953	0.063	4	310	690	--	--	--	--	0.70	900+	18	25.6
G5	M953	0.063	4	324	820	--	--	--	--	--	900+	--	--
H240	M953	0.063	4	180	720	--	--	--	--	0.60	900+	--	--
H236	M953	0.063	4	255	735	--	--	--	--	0.60	900+	--	--

Table 9 Creep properties of N-155 alloy sheet at 2000 F (1093 C)

Specimen No.	Heat No.	Gage (ins.)	Creep Stress (Ksi)	Time to Indicated Plastic Creep Strain (sec)						Plastic Creep in 900 sec (%)	Time to Rupture (sec)	Elong. (%)	Residual UTS at 2000°F (Ksi)
				0.2%	0.5%	1%	2%	3%	4%				
M628	M5171	0.063	10	2.0	5.1	10	18	26	30	--	102	--	--
M653	M5171	0.063	10	1.3	2.7	4.8	10	16	21	--	30	--	--
Q38	M5148	0.063	10	1.5	3.0	6.0	12	18	24	--	70	19	--
Q39	M5148	0.063	10	2.2	5.0	8.9	16	25	34	--	121	20	--
Q80	M5148	0.063	10	1.2	3.5	6.0	11	15	20	--	30	14	--
Q81	M5148	0.063	10	1.2	3.0	5.0	10	15	19	--	31	19	--
M631	M5171	0.063	8	3.3	9.0	20	42	65	84	--	408	--	--
M648	M5171	0.063	8	2.9	6.8	17	30	44	54	--	120	--	--
Q40	M5148	0.063	8	5.0	11	21	42	61	81	--	256	18	--
Q41	M5148	0.063	8	5.0	13	23	53	80	107	--	321	18	--
Q42	M5148	0.063	8	6.0	15	30	58	87	121	--	280	16	--
M62*	M1675	0.063	7	11	34	66	120	--	--	--	120+	15	18.1
M63*	M1675	0.063	7	11	31	63	132	--	--	--	132+	15	21.2
M251	M1492	0.078	7	5.5	15	35	72	--	--	--	72+	36	21.2
M250	M1492	0.078	7	9.0	27	54	102	--	--	--	102+	30	22.2
K554	M5015	0.078	7	12	45	85	168	--	--	--	168+	--	--
M248	M1492	0.078	7	--	16	41	78	--	--	--	78+	--	21.2
Q43	M5148	0.063	7	5.5	13	26	51	75	104	--	534	18	--
Q44	M5148	0.063	7	6.0	14	29	60	88	117	--	393	15	--
Q46	M5148	0.063	7	5.5	18	38	80	124	170	--	521	18	--
Q70	M5148	0.063	7	7.0	20	42	79	125	169	--	382	15	--
Q71	M5148	0.063	7	7.0	20	42	75	112	138	--	339	15	--
K695	M1475	0.063	6	21	73	145	305	--	--	--	305+	14	17.7
K697	M1475	0.063	6	28	81	182	312	--	--	--	312+	16	19.1
M228	M1492	0.078	6	9.1	32	72	168	--	--	--	168+	24	20.2
M229	M1492	0.078	6	26	61	120	192	--	--	--	192+	24	20.4
M64*	M1675	0.063	6	15	43	91	186	--	--	--	186+	15	19.6
M66*	M1675	0.063	6	28	62	115	204	--	--	--	204+	13	21.5
K557	M1675	0.063	6	19	50	98	210	--	--	--	450+	--	--
Q48	M5148	0.063	6	13	37	91	195	313	409	--	900+	14	--
Q73	M5148	0.063	6	15	36	75	133	210	291	--	744	12	--
K700	M1475	0.063	5	57	159	355	630	--	--	--	630+	17	18.7
M231	M1492	0.078	5	28	77	170	318	--	--	--	318+	24	19.7
M235	M1492	0.078	5	47	120	305	480	--	--	--	480+	27	21.5
M236	M1492	0.078	5	56	125	330	606	--	--	--	606+	23	21.2
M68*	M1675	0.063	5	35	78	170	288	--	--	--	288+	15	18.4
K560	M5015	0.078	5	65	160	315	555	--	--	--	900+	--	--
Q49	M5148	0.063	5	85	155	245	360	485	615	--	900+	23	15.6
Q51	M5148	0.063	5	28	82	190	340	515	694	--	900+	25	17.2
K701	M1475	0.063	4	96	340	600	--	--	--	1.6	900+	17	18.2
K704	M1475	0.063	4	135	332	562	--	--	--	1.8	900+	15	16.7
M233	M1492	0.078	4	99	308	555	--	--	--	1.6	900+	28	21.4
M70*	M1675	0.063	4	85	240	430	762	--	--	--	762+	22	19.8
M72*	M1675	0.063	4	82	240	415	708	--	--	--	708+	21	19.2
K561	M1475	0.063	4	140	300	600	--	--	--	1.6	900+	--	--
Q52	M5148	0.063	4	100	260	493	818	--	--	2.2	900+	25	18.7
Q53	M5148	0.063	4	90	268	570	--	--	--	1.7	900+	25	19.3
Q54	M5148	0.063	4	120	295	510	815	--	--	2.4	900+	25	16.9
K556	M5015	0.078	4	100	300	525	--	--	--	1.8	900+	--	--

* Transverse specimens

Table 10 Creep properties of N-155 alloy sheet at 2100 F (1149 C)

Specimen No.	Heat No.	Gage (ins.)	Creep Stress (Ksi)	Time to Indicated Plastic Creep Strain (sec)						Plastic Creep in 900 sec (%)	Time to Rupture (sec)	Elong. (%)	Residual UTS at 2100°F (Ksi)
				0.2%	0.5%	1%	2%	3%	4%				
M655	M5171	0.063	8	1.0	2.5	5.4	11	16	20	--	84	23	--
M656	M5171	0.063	8	1.5	3.6	7.2	14	20	26	--	72	20	--
M657	M5171	0.063	8	1.0	2.0	4.0	8	11	14	--	50	22	--
M673	M5171	0.063	8	1.0	2.8	5.3	11	16	21	--	97	23	--
M674	M5171	0.063	8	1.0	2.0	3.9	8.0	11	16	--	--	--	--
Q19	M5148	0.063	8	0.6	1.4	2.6	5.2	8.0	11	--	50	19	--
Q20	M5148	0.063	8	1.0	2.6	5.0	10	15	20	--	88	20	--
Q21	M5148	0.063	8	0.8	1.8	3.4	6.6	8.6	11	--	51	20	--
M658	M5171	0.063	7	1.5	3.3	7.2	14	20	26	--	96	24	--
M664	M5171	0.063	7	1.0	3.5	7.5	15	23	30	--	81	16	--
M665	M5171	0.063	7	0.6	2.6	5.5	12	19	26	--	93	17	--
M670	M5171	0.063	7	1.0	2.7	5.2	11	16	22	--	90	19	--
M671	M5171	0.063	7	1.0	3.0	5.5	11	17	23	--	105	16	--
Q22	M5148	0.063	7	1.4	3.5	6.8	11	17	23	--	122	20	--
Q23	M5148	0.063	7	1.9	4.1	8.0	13	19	26	--	122	20	--
Q24	M5148	0.063	7	1.3	3.0	6.0	11	15	20	--	70+	--	--
M646	M5171	0.063	6	3.0	7.2	18	33	47	60	--	192	16	--
M647	M5171	0.063	6	3.3	8.6	18	30	49	66	--	156	20	--
M662	M5171	0.063	6	4.0	10	22	46	66	84	--	306	18	--
M668	M5171	0.063	6	2.0	5.7	11	24	37	50	--	255	16	--
M672	M5171	0.063	6	3.0	7.0	14	29	45	61	--	270	17	--
Q25	M5148	0.063	6	--	--	17	33	50	65	--	208	17	--
Q27	M5148	0.063	6	3.0	7.5	14	30	45	60	--	285	20	--
M634	M5171	0.063	5	3.9	11	30	66	90	120	--	396	14	--
M636	M5171	0.063	5	4.1	11	30	60	90	120	--	306	13	--
M663	M5171	0.063	5	9.5	24	51	114	168	238	--	565	13	--
M666	M5171	0.063	5	3.5	14	28	58	85	105	--	399	16	--
M667	M5171	0.063	5	3.5	9.5	21	45	75	107	--	360	15	--
Q28	M5148	0.063	5	12	--	65	115	170	227	--	537	12	--
Q30	M5148	0.063	5	8.0	22	42	84	100	115	--	275	15	--
M637	M5171	0.063	4	9.5	30	78	150	210	288	--	678	12	--
M639	M5171	0.063	4	9.0	25	66	156	265	336	--	846	12	--
M640	M5171	0.063	4	12	42	96	198	300	402	--	870	10	--
Q81	M5148	0.063	4	25	70	145	291	435	603	--	900+	17	10.6
Q82	M5148	0.063	4	--	46	102	209	335	475	--	900+	15	12.1
Q83	M5148	0.063	4	11	36	76	--	--	--	--	--	--	--
Q83	M5148	0.063	4	30	60	101	165	244	322	--	535	12	--
Q84	M5148	0.063	4	16	41	82	190	260	--	--	--	--	--
Q85	M5148	0.063	4	11	40	95	150	305	--	--	--	--	--
Q87	M5148	0.063	4	15	50	111	262	416	565	--	900+	20	11.8
M641	M5171	0.063	3	--	110	276	522	750	--	--	900+	16	--
M643	M5171	0.063	3	38	105	258	528	756	--	--	900+	16	--
M644	M5171	0.063	3	36	105	--	564	880	--	--	900+	16	--
M642	M5171	0.063	3	30	115	258	456	636	--	--	--	--	--
Q85	M5148	0.063	3	40	108	280	--	--	--	--	900+	25	12.2
Q86	M5148	0.063	3	50	123	280	570	--	--	--	900+	32	--

Table 11 Comparative tensile and creep properties of N-155 alloy sheet

	Property	Rate	Temperature 1600°F (871°C)	1800°F (982°C)	1900°F (1038°C)	2000°F (1093°C)	2100°F (1149°C)
Tensile Properties	Tensile proportional limit stress	0.0001 in./in./sec	20,500	9,000	5,200	3,000	2,000
		0.001 in./in./sec	21,300	14,000	10,000	7,000	5,000
		0.01 in./in./sec	21,300	16,700	13,400	10,000	7,500
	0.2% yield strength	0.0001 in./in./sec	27,000	12,500	8,000	5,000	3,000
		0.001 in./in./sec	29,500	17,000	11,500	8,000	6,000
		0.01 in./in./sec	30,000	25,500	18,700	14,000	11,000
	Ultimate strength	0.01 in./in./sec	44,000	25,500	18,700	14,000	11,000
		0.1 in./in./sec	32,500	35,000	27,600	21,500	16,000
Creep Properties	Stress for 1% plastic creep	in 1 sec	41,100	23,500	17,600	12,700	9,200
		10 sec	33,000	18,500	12,000	8,700	6,000
		60 sec	25,600	12,900	8,200	6,100	4,200
		600 sec	19,000	8,000	5,000	3,800	2,600
		900 sec	18,000	7,400	4,700	3,400	2,350
	Stress for 4% plastic creep	in 1 sec	43,400	24,600	18,400	13,700	10,300
		10 sec	38,600	21,700	15,300	11,000	8,000
		60 sec	30,800	17,300	10,900	8,000	5,600
		600 sec	23,200	10,500	6,800	5,100	3,700
		900 sec	22,000	9,500	6,200	4,400	3,300
	Stress for indicated plastic creep rate	0.001 in./in./sec	33,000	18,500	12,000	8,700	6,000
		0.0001 in./in./sec	24,000	11,000	7,200	5,200	3,300
	Stress for indicated life to rupture	60 sec	36,000	21,000	15,000	9,500	7,500
		600 sec	27,000	13,500	9,000	6,500	4,000

NOTE: Data Condensed from Tables VI to X and Figures 2, 3, and 12 to 16.

relatively uniform at all temperatures from room temperature to 2100 F. The typical values shown in Figs. 2, 3, 4, and 5 are therefore considered suitable for use as design-allowable strengths.

2 The tensile, yield, and ultimate strengths of N-155 alloy are strain-rate sensitive. Rate effects are noted at critical temperatures which vary with the strain rate used.

3 The tensile elongation and Young's modulus of N-155 alloy are both relatively insensitive to changes in strain rate.

4 The speeds used in conventional room-temperature tensile tests (about $^1/_2$ min to yield, then about $1^1/_2$ min to fracture) are too slow for determining the realistic short-time tensile strengths at high temperatures due to the major effects of creep.

5 Prior plastic-creep deformation of up to 4 per cent in 900 seconds has no significant effect on the residual tensile strength of annealed N-155 alloy sheet.

6 Creep rates determined at high temperatures and high stresses are extremely rapid. Where such severe temperatures and high, sustained, load levels are encountered in service, the design must be based on creep strength rather than on tensile strength even where comparatively short-time service life is anticipated.

Future Work

The data reported here will be analyzed using time-temperature-parameter methods to determine possible correlations with long-time creep properties. A program to determine the compression strengths and the short-time compression-creep properties of N-155 alloy is also being undertaken. This type of data would be of great value in aircraft and missile design since compression loading very often becomes the governing design criteria in these applications.

Acknowledgments

The author wishes to acknowledge the help received from the many people involved in this work. Particular thanks are extended to A. V. Levy for the many suggestions and critical review of this paper and to C. A. Drury and A. F. Konigsfeld who carried out the tests. Thanks are also extended to the Marquardt Corporation for making this research possible and for permission to publish this paper.

References

1 R. H. Sparling, "Testing in the Guided-Missile Industry," *ASTM Bulletin*, December, 1956.

2 W. H. Steuer, "Metals for High-Speed Flight," *Metals Progress*, April, 1957.

3 S. E. Bramer, K. G. Kahmann, and R. Titus, "A Programming, Universal, Elevated-Temperature-Testing Machine," The Marquardt Corporation, Van Nuys, Calif.

4 C. L. Dotsen and J. R. Kattus, "Tensile Properties of Aircraft Structural Metals at Various Rates of Loading After Rapid Heating," WADC TR 55-199.

DISCUSSION

J. R. Kattus[3]

The author has certainly done an excellent and thorough job of evaluating the load-carrying capacity of annealed N-155 alloy sheet under conditions of rapid heating, rapid loading, and short times at temperature. The effects of these conditions on the mechanical properties of N-155 alloy are, in general, quite similar to the effects that we have found in our evaluations of the short-time tensile and creep properties of a large number of other alloys. For example, we have found that the relative effects of tensile strain rate become quite large in all alloys when the temperature is increased above a certain level. This temperature corresponds roughly to the recrystallization temperature of each alloy, strength increasing significantly with increasing strain rate at higher temperatures as shown in Figs. 2 and 3. In contrast to the author's data for N-155, we have found that the tensile

[3] Head of Metallurgy Division, Southern Research Institute, Birmingham, Ala.

strength of most alloys also tends to increase, but to a relatively much smaller degree, with increasing strain rate at lower temperatures. In some alloys, however, over certain temperature ranges, strain-rate-sensitive structural changes such as strain aging reduce or completely counterbalance the inherent strengthening effect of increasing strain rate.

The finding that variations in holding time at temperature from 2 to 30 min have no significant effect on the tensile properties of annealed N-155 alloy should not be construed to mean that this is not an important variable in all alloys. The annealed N-155 happens to be a stable material over the entire range of test temperatures. Many other materials, particularly those that are hardened by heat-treatment or by cold work, undergo changes in metallurgical structure at high temperatures. These structural changes are often accompanied by a marked deterioration in mechanical properties. Since the structural changes are dependent upon both time and temperature, a small variation in holding time at certain temperatures can have quite significant effects on strength properties. In some comparative evaluations of cold-worked and annealed N-155 alloy, for example, we have found that the cold-worked material retained a constant relative superiority in strength over the annealed material at temperatures up to 1500 F regardless of variations in holding time from 10 sec to 30 min. At higher temperatures, the recrystallization process in the cold-worked material caused a decrease in strength with increasing holding time, whereas the annealed material was unaffected by variations in holding time. After a 10-sec holding time at 2000 F, for example, the cold-worked material retained the same relative strength advantage that it had had at lower temperature; whereas after a 30-min holding time the strength of the cold-worked material had decreased to a level below that of the annealed material. The length of time that the structural material in a particular application is to be held at elevated temperature as well as the temperature level and the strain rate can have an important bearing on the choice of a material for that application.

D. M. Lorimer[4] and M. Aarnes[4]

For designs at elevated temperature it is apparent that, in order to reduce weight, materials have to be used past the 0.2 per cent yield strength. Bernett's paper tends to point out that the yield strength at elevated temperatures is affected by strain rate. Below a certain strain rate at a given temperature the material creeps during test. A boundary should be selected in terms of stress, strain, time, and temperature below which the conventional methods of strength checking analysis will be used and above which the creep deformation has to be considered.

The influence of strain rate can be shown by examining Fig. 5 and considering 1800 F data only. If loading was purely elastic to a given strain, Young's modulus E and the secant modulus E_s would be equal. The variance of E_s from E is a measure of creep during loading. Selecting 2 mils/in. strain we have

$$E_s = \sigma_s/\epsilon = \frac{22{,}500}{0.002} = 11.25 \times 10^6 \quad (\dot{\epsilon} = 0.01 \text{ in/in/sec})$$

$$E_s = \sigma_s/\epsilon = \frac{17{,}000}{0.002} = 8.5 \times 10^6 \quad (\dot{\epsilon} = 0.001 \text{ in/in/sec})$$

$$E_s = \sigma_s/\epsilon = \frac{12{,}900}{0.002} = 6.45 \times 10^6 \quad (\dot{\epsilon} = 0.0001 \text{ in/in/sec})$$

From Fig. 4, we have E at 1800 F as 13×10^6, so:

$$\frac{E_s}{E} = \frac{11.25 \times 10^6}{13 \times 10^6} \times 100 = 86.5\%$$

$$\frac{E_s}{E} = \frac{8.5 \times 10^6}{13 \times 10^6} \times 100 = 65.3\%$$

$$\frac{E_s}{E} = \frac{6.45 \times 10^6}{13 \times 10^6} \times 100 = 49.6\%$$

so that for:

$\dot{\epsilon} = 0.01$ in/in/sec appr. 13.5 per cent of the strain is creep
$\dot{\epsilon} = 0.001$ in/in/sec appr. 34.7 per cent of the strain is creep
$\dot{\epsilon} = 0.0001$ in/in/sec appr. 50.4 per cent of the strain is creep

if the small amount of plastic deformation is ignored. This shows that Bernett's strain rate of 0.1 in/in/sec for loading in creep tests is mandatory to approach elastic loading.

The use of I^2R heating can be criticized because it introduces hot spots in the specimen and will affect the results. In addition, if any microcracks open up during plastic strain, I^2R heating has a tendency to weld the ends of these microcracks thus reducing notch effect and giving high readings. Radiant heating upon stressing would be more realistic in aircraft and missile environments.

The recording of true stress-strain in short time creep testing is desirable. There is a difference between constant load and constant stress creep (refer to Andrade's work). Almost all creep work could be improved in this direction.

This work is a significant step in the defining of metals' properties in the short time creep and tensile range.

Warren K. Smith[5]

The author has reported on a well planned and interesting investigation of short-time high-temperature properties of an important alloy. It is gratifying to the writer, after having shared similar misgivings about the use made of high-temperature data, to see these timely remarks and cautions directed to the design engineer.

Since the annealed N-155 alloy is quite stable for the holding times used in these tests, it is of no consequence that a 15-min hold was used for the tensile test and a 60-sec hold for the creep test. However, if this is regular practice for all materials, an awkward variable will be present in comparing tensile and creep tests on alloys with unstable microstructure. The greater strength of a heat-treated or cold-worked alloy can be used to advantage in a short-time application and is in many missile parts. Therefore, it would seem logical that no hold time at all should be specified in either test. Resistance heating of the specimens brings them quite uniformly up to test temperature, regardless of the heating rate. For those situations where, for example, a missile is carried in a supersonic airstream and heated for a while before firing, a special holding time simulation had better be prescribed rather than attempt to make it into a universal standardized test.

To carry the discussion of service simulation further, one might ask if the elevated-temperature tensile test above the creep point has any direct application at all. A service condition consisting of a single loading application at a certain strain rate and no holding time would be required. There are very few examples of such service conditions to be found. In nearly every case the load must be sustained for some time interval, even though it is only a few seconds. The author has been careful to point out repeatedly that creep test data should be used if there is any chance that the load may remain even for a second or two at high temperature and stress. It would seem, therefore, that the efforts expended on elevated-temperature tensile testing should instead be on short-time creep tests down to temperatures and stresses that will result in practically no measurable plastic

[4] Boeing Airplane Company, Seattle, Wash.

[5] Code 4061, Aero. Struct. Matls. Res. Eng., U. S. Naval Ordnance Test Station, China Lake, Calif.

deformation within some reasonable time limit. Below this temperature level, the "transition temperature" in this paper, a tensile test is proper. The writer believes it would be easier to determine this transition temperature by short-time creep tests than by doing both high and low strain-rate tensile tests in the majority of laboratories.

Author's Closure

A number of important aspects related to the short time mechanical behavior of materials were stressed in the discussions. This will help greatly in creating a better understanding of the problems involved in testing of materials and those encountered in using the test data in design.

Regarding the comments on hold time at temperature made by Mr. Kattus and Mr. Smith, it was most certainly not the author's intent to create any impression that this variable is unimportant. Longer hold-times can result in much higher or much lower measured mechanical properties depending on the particular material and test temperatures involved. Furthermore, a hold time of "zero" or any other single time interval is not going to be universally acceptable either. The designer and the materials engineer must continue to analyze each design jointly. The final configurations of the high-temperature components must be based on materials data determined at critical temperatures and times which closely simulate the conditions encountered in service.

Elevated temperature testing is expensive and time consuming; however, there are no alternatives at present. The activity in the high-temperature testing field is increasing constantly and more and more data are becoming available from many different laboratories. In this regard, it is strongly urged that greater care be taken in accurately defining the conditions under which the test data were obtained. This supplementary information must be known before an intelligent appraisal of any reported data can be made.

Mr. Smith questions the relative value of tensile tests carried out at temperatures above the "creep point." The author agrees that, in the majority of cases, creep type data are much more useful for actual design purposes. Where ultrashort-time transient loads which are much higher than the steady-state loads are encountered, however, rapid strain rate tensile data could be used to advantage. Thus it is essential that both creep and tensile data for each high-temperature material are available to the design engineer.

Lorimer and Aarnes raised some objections to the use of electrical resistance heating in the determination of mechanical properties. For the rapid heating rates and high temperatures required, resistance is the only practical method available today. A limited amount of "rapid rate" data generated on equipment using standard furnaces, radiant heaters, hot flowing fluids, etc., have been reported in the literature. It is obvious that in each case difficulties were encountered either in obtaining the higher test temperatures or in adapting suitable strain measuring systems to the equipment. The data that have been obtained are therefore sketchy. In any case, wherever a direct comparison with our resistance heating data has been possible, the agreement is excellent. In addition to the spot checks on N-155 alloy sheet data, comparisons have also been obtained on 2024 aluminum, HM21 magnesium, 4130 low alloy steel, 6Al-4V titanium, 17–7 stainless steel, and Rene 41 nickel base alloy. For the purpose of mechanical property determinations therefore, it is immaterial which method of heating is used. Comparable properties will be obtained provided that both the heating rate and hold time at temperature in each case are equivalent.

A Photoelastic Analysis of Machining Stresses[1]

E. USUI
H. TAKEYAMA
Research Engineers, Government Mechanical Laboratory, Tokyo, Japan

Direct measurements of the distributions of normal and frictional stresses on a rake face under cutting conditions have been considered to be practically impossible. However, as reported in this paper, the stress distributions have been successfully obtained photoelastically by using a tool made of a photoelastic material.

According to the authors' experiment, the frictional stress on the rake face is distributed uniformly over a wide range of the tool-chip contact length, but it decreases rapidly near the point of chip-separation on the rake face. As to the normal stress, it has a peak near the cutting edge, being rather stationary in the middle part of the contact length and decreasing gradually toward the point of chip-separation.

Introduction

VARIOUS THEORIES of orthogonal cutting have been developed hitherto, but in most of the theories the so-called coefficient of friction on a rake face which is calculated from the equation $\mu = \dfrac{F_H \sin\alpha + F_V \cos\alpha}{F_H \cos\alpha - F_V \sin\alpha}$ by using the two components of the cutting force measured with a tool dynamometer is assumed. However, it must be noticed that this equation involves two important assumptions implicitly. Namely, the coefficient of friction is assumed to be constant everywhere over the tool-chip interface based upon Coulomb's law and the distributions of the interface stresses are assumed implicitly to be uniform. It is quite doubtful whether the simple Coulomb's law does hold under such a severe condition as at the tool-chip interface in metal machining, and also the second assumption will be meaningless if a peak force exists at the cutting edge as reported already in the authors' previous paper [1].[2] Therefore, a precise information of the stress distributions on a rake face must be one of the most essential problems for a more reliable analysis of the mechanism of metal machining as well as for the better practice.

Photoelastic analyses of metal machining have been conducted by several persons, but in most cases [2, 3] the stresses within the photoelastic materials used as the materials cut have been studied. However, by such a photoelastic technique an analysis of the stress zone ahead of the rake face is almost impossible because of its plastic deformation. In the authors' experiment, on the contrary, a photoelastic material is used as the cutting tool so that the normal and frictional stresses on the rake face can be detected.

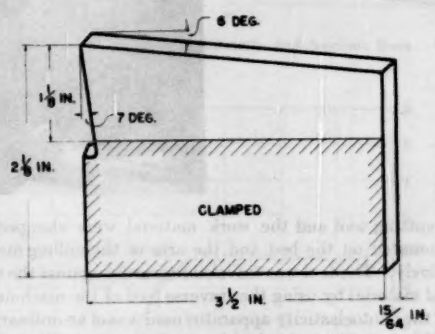


Fig. 1 General view of photoelastic cutting tool

Test Procedure

A lead plate was machined two-dimensionally with a tool made of a photoelastic material, epoxy resin, which was specially designed, as shown in Fig. 1, to be strong enough for the cutting force and also to avoid the boundary effect of the tool, especially, the influence of its clamping. The reason why lead was selected as the material cut was that it is soft enough for the tool of epoxy resin. The width of the plate which always corresponded to the width of cut was 0.2 in. The depth of cut was 0.0342 in. and the cutting speed was 0.71 in. per minute. The tool was ground with a cutter-grinder to have the shape of (7, 0, 6, 0, 0, 0, 0) and the rake-face roughness of 0.000173 in. A milling machine of knee type was used for the orthogonal machining test. The photo-

[1] Contributed at the Annual Meeting, Hiroshima city, Hiroshima prefecture, Japan, April 15, 1958, of The Society for Precision Mechanics of Japan.

[2] Numbers in brackets designate References at end of paper.

Contributed by the Production Engineering Division and presented at the Annual Meeting, Atlantic City, N. J., November 29–December 4, 1959, of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS.

NOTE: Statements and opinions advanced in papers are to be understood as individual expressions of their authors and not those of the Society. Manuscript received at ASME Headquarters, July 14, 1958. Paper No. 59—A-37.

Nomenclature

μ = mean coefficient of friction on rake face
α = rake angle
F_H = principal component of cutting force
F_V = feed component of cutting force
σ_1 = algebraically larger principal stress
σ_2 = algebraically smaller principal stress
ϕ = angle between direction of σ_1 and x-axis
σ_y = normal stress in direction of y-axis
σ_{y0} = value of σ_y when $y = y_0$
S = distance from cutting edge to acting point of resultant cutting force on rake face
σ = normal stress on rake face
τ = frictional stress on rake face
θ = angle between direction of resultant cutting force and rake face
σ_s = normal stress in shear plane

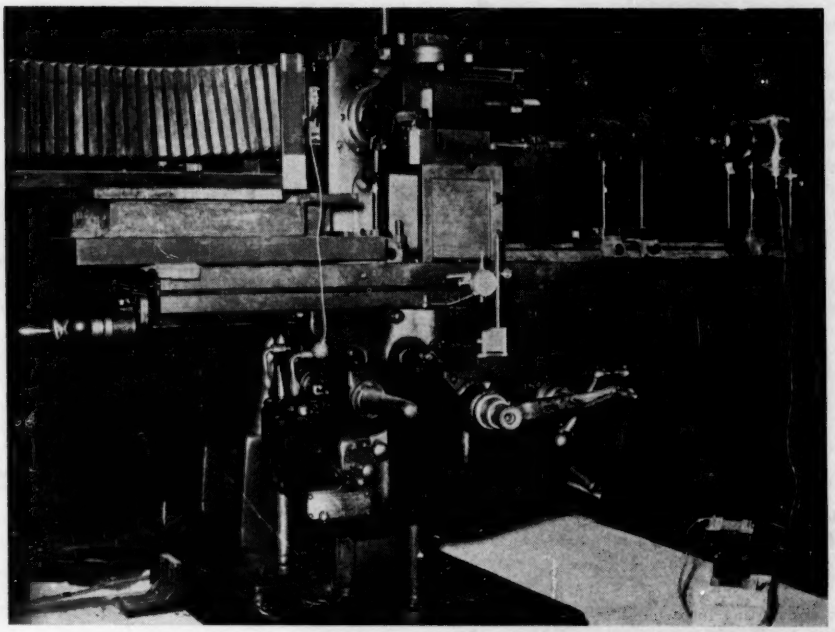
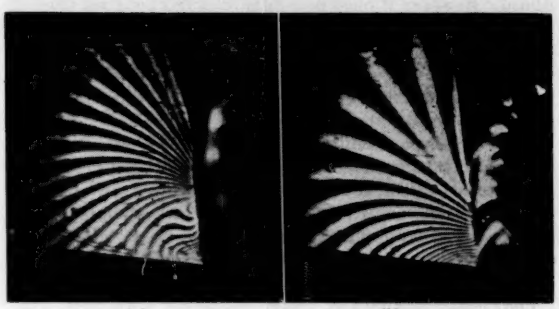


Fig. 2 Setup for photoelastic experiment

elastic cutting tool and the work material were clamped with vices mounted on the bed and the arm of the milling machine, respectively. The tool was traveled forwardly against the rigidly clamped material by using the traverse feed of the machine. Although the photoelasticity apparatus used was of an ordinary type for the fringe method, the field lenses were not used because the stress zone of the cutting tool was small enough to be covered by the nearly parallel light provided by two condenser lenses as shown in Fig. 2. The apparatus was arranged on the table of the milling machine so as to be moved together with the photoelastic tool. Thus the internal stresses of the tool could be clearly observed during cutting. The isochromatic and isoclinic lines were unexpectedly stationary during machining, so that the photographing was rather easy fortunately. The vibration of the milling machine did not affect the photographs because the cutting speed was low, and both the tool and the whole optical system including the camera were mounted on the same table.

Isochromatic and Isoclinic Lines

Fig. 3 shows two examples of the fringe patterns. The photograph (*a*) is the isochromatic pattern while cutting and (*b*) shows the pattern when the tool was stopped and then statically pressed against the readymade chip. Apparently the two fringe patterns are quite different from each other. The fringes in (*b*) are rather regularly spaced, converging together gradually toward the cutting edge, whereas the pattern in (*a*) is more complex; that is to say, the fringes are closely concentrated both at the cutting edge and near the point of chip-separation on the rake face, but they are rather sparse in the middle region. Each fringe corresponds to the line upon which the difference of the two principal stresses $\sigma_1 - \sigma_2$ has the same value. Therefore the distribution of the fringe corresponds to that of $\sigma_1 - \sigma_2$. $\sigma_1 - \sigma_2$ is also related directly with the shear stress at each point as will be mentioned later. Thus it can be concluded that the distribution of the frictional stress on the rake face obtained by a statical experiment will never reveal the real situation during actual machining. Nevertheless several persons have carried out hitherto such statical experiments, in each of which a photoelastic tool was statically pressed against a specially designed



(a) (b)

Fig. 3 Dynamic (a) and static (b) fringe patterns

model [3] or a readymade chip forced with a steel tool [2].

Fig. 4 shows an example of the isoclinic lines photographed while cutting. Although the isoclinic lines are usually traced or photographed by using other photoelastic material of less sensibility than that for the fringe patterns, in the experiment the lines were obtained by using the same tool immediately after taking the photograph of the fringe pattern, because all experimental conditions should be kept the same in both the tests for the fringe pattern and the isoclinic lines.

Fig. 5 shows the composite pattern of the isochromatic and isoclinic lines. The isochromatic lines are denoted by the thin lines and the numbers put to them represent the respective fringe orders. The fringe of zero order was identified by the coloration of each fringe observed when the filter was withdrawn. The isoclinic lines are shown by the bold lines. Taking tension positive and compression negative, the angle between the direction of algebraically larger principal stress σ_1 and one axis of the polarized light OP_1 will be given for each isoclinic line as indicated in Fig. 5.

Frictional Stress on Rake Face

If the difference between the two principal stresses $\sigma_1 - \sigma_2$ and the angle ϕ between the direction of algebraically larger principal

Fig. 4 Isoclinic line of 40 deg, rake angle = 7°. Material cut lead; tool material epoxy resin; tool shape 7, 0, 6, 0, 0, 0, 0; width of cut 0.2 in.; depth of cut 0.0342 in.; cutting speed 0.71 ipm; cutting fluid dry.

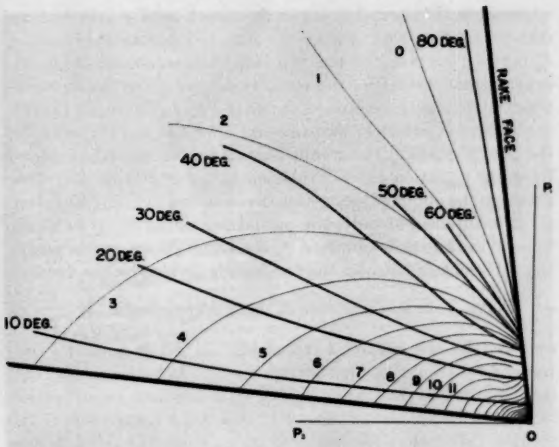


Fig. 5 Composite pattern of isochromatic and isoclinic lines. Cutting conditions same as Fig. 4.

Table 1 Calculating data of frictional stress on rake face

Distance from cutting edge, in.	ϕ, deg	sin 2ϕ	$\sigma_1 - \sigma_2$, f.o.	τ, f.o.
0.008	14.2	0.480	21.0	5.04
0.013	15.4	0.515	19.0	4.90
0.020	17.0	0.559	17.0	4.75
0.040	23.5	0.750	13.0	4.88
0.049	27.0	0.809	12.0	4.85
0.059	29.7	0.875	11.0	4.81
0.067	37.0	0.961	9.8	4.71
0.075	40.3	0.990	9.0	4.45
0.084	47.0	0.998	8.0	4.00
0.093	52.7	0.970	4.8	2.33
0.100	57.0	0.914	3.0	1.37
0.107	67.0	0.719	2.0	0.72
0.113	77.0	0.438	1.0	0.22

stress and x-axis are given, the shear stress can be obtained by the following equation

$$(\sigma_1 - \sigma_2) \sin 2\phi = 2\tau \tag{1}$$

where σ_1 and σ_2 are algebraically larger and smaller principal stress, respectively, and τ is the shear stress. Here, $\sigma_1 - \sigma_2$ is proportional to the fringe order shown in Fig. 5. Taking the direction of the rake face to be x-axis, ϕ can be obtained by translating the known angle between σ_1 and OP_1 into the angle between σ_1 and the x-axis. The values of ϕ and $\sigma_1 - \sigma_2$ at various points on the rake face are summarized in Table 1 together with the values of sin 2ϕ and calculated τ.

Fig. 6 shows the distribution of thus obtained frictional stress on the rake face. It is observed that the frictional stress is distributed almost uniformly over three quarters of the tool-chip contact length from the cutting edge as reported already in the authors' previous paper [1] and the value of the stress seems to be nearly equal to the shear strength of the material cut for the reason mentioned later. The frictional stress, however, decreases rapidly near the point of chip-separation on the rake face. Although it was reported in the authors' previous paper [1] that the distribution is uniform all over the tool-chip contact area, this is not necessarily a rough approximation considering the steepness of the distribution curve near the point of chip-separation, the contact of this area not being taken into consideration in the previous paper. As a matter of fact, the actual cutting process must be governed by the region of uniform stress, and the boundary near the point of chip-separation plays only a minor role in machining. As indicated by the dotted line in Fig. 6, the frictional stress at the cutting edge seems to be high.

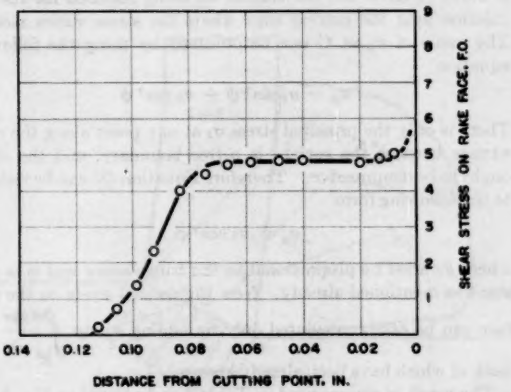


Fig. 6 Distribution of shear stress on rake face. Cutting conditions same as Fig. 4.

The angle ϕ at any point on the relief should have a definite value because the relief is a free boundary where the principal stresses have the same direction. Therefore, theoretically speaking, the angle at the cutting edge will have the same value. On the contrary, the fringe order rises up steeply toward the cutting edge as observed in Fig. 5. Such trends make the frictional stress near the cutting edge steeply high theoretically due to Equation (1). Considering that the existence of the singular force at the cutting edge was proved experimentally in the authors' previous paper [1], the peak frictional stress may be explicable as the component of the singular force in the tangential direction of the rake face. However, from a practical point of view, the relief surface near the cutting edge will be touched and compressed by the machined surface instead of being free. Therefore the stress field near the cutting edge must be more complex and so any assertive explanation about it cannot be given at the present stage.

Normal Stress on Rake Face

One of the two equations for an equilibrium condition in an elastic body can be written in integral form as follows

$$\sigma_y = \sigma_{y0} - \int_{y_0}^{y} \frac{\partial \tau}{\partial x}\, dy \tag{2}$$

where σ_y is the normal stress in the direction of y-axis, and σ_{y0} is the value of σ_y when y is y_0. In order to calculate the values of σ_y on the rake face, the tool is divided into a number of quadrilaterals small enough for the accurate calculation by the lines perpendicular to the rake face, as shown in Fig. 7. Assuming the distance between arbitrary two adjacent lines AD and CE to be dx, the normal stress at the middle point P on DE can be computed by using Equation (2) as explained hereafter. Since the shear stress can be obtained along each dividing line by using Equation (1) as mentioned already, the partial derivative $\partial\tau/\partial x$ in Equation (2) can be easily obtained if, at a certain point on the y-axis, the difference $d\tau$ between the two lines is divided by the distance dx. Table 2 shows the values of the definite integral $\int_{y_0}^{y} \frac{\partial\tau}{\partial x}\, dy$ obtained by the graphical integrating, in which y_0 and y correspond to Q and P in Fig. 7, respectively. As to the value of σ_{y0}, it should be the value at Q, strictly speaking, but its calculation is practically complex or difficult. This is the reason why in the present calculation the value of σ_y at C on the relief (a free boundary) is substituted for σ_{y0}. In this case, of course, the smaller dx, the better the approximation. In this calculation dx is 0.013 to 0.005 in., the smaller dx being required for the calculation near the cutting edge where the stress varies steeply. The value of σ_y at C can be obtained by using the following equation

$$\sigma_y = \sigma_1 \sin^2 \phi + \sigma_2 \cos^2 \phi \qquad (3)$$

There is only the principal stress σ_2 at any point along the relief surface because the surface is a free boundary, and the stress ought to be compressive. Therefore Equation (3) can be reduced to the following form

$$\sigma_y = \sigma_2 \cos^2 \phi \qquad (4)$$

where σ_2 must be proportional to the fringe order and ϕ is constant as mentioned already. Now the normal stress on the rake face can be easily calculated only by adding σ_{y0} to $\int_{y_0}^{y} \frac{\partial\tau}{\partial x}\, dy$, both of which have been already known.

The result is summarized in Table 2 and plotted in Fig. 8.

Table 2 Calculating data of normal stress on rake face

Distance from cutting edge, in.	σ_{y0}, f.o.	$\int_{y_0}^{y} \frac{\partial\tau}{\partial x}\, dy$, f.o.	σ_y, f.o.
0.011	14.23	0.21	14.02
0.016	12.35	−0.48	12.83
0.020	12.35	−0.28	12.63
0.033	9.35	−3.03	12.38
0.047	7.40	−3.96	11.36
0.060	6.01	−2.71	8.72
0.073	5.12	−1.80	6.92
0.087	4.37	1.23	3.14
0.100	3.80	3.07	0.73
0.113	3.33	3.23	0.10

It is worthy of attention that the normal stress near the cutting edge is extraordinarily high, or there exists a peak force at the cutting edge. Such a trend was also proved by the authors' experiment as reported in the previous paper [1]. Another part than the aforementioned special region can be divided into two parts from the viewpoint of stress distribution. In the region adjacent to the so-called special region, the distribution of the normal stress is rather stationary, while in the outermost region the stress decreases almost linearly to vanish at the point of chip-separation. It is very interesting to compare the distributions of both frictional and normal stresses on the rake face. While the former is distributed uniformly in a wide range of the tool-chip contact length, the latter varies independently of the former. This

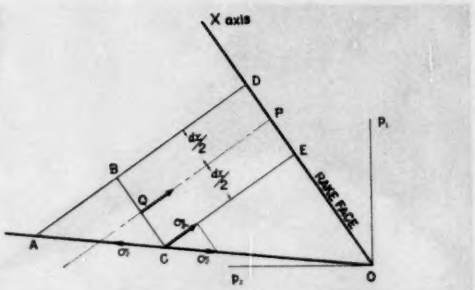


Fig. 7 Calculating procedure of normal stress on rake face

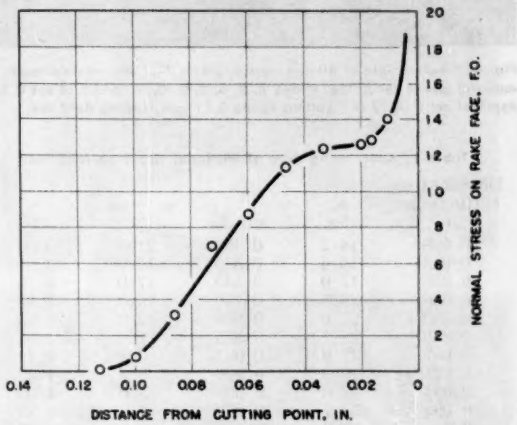


Fig. 8 Distribution of normal stress on rake face. Cutting conditions same as Fig. 4.

seems to mean that the frictional stress increases to a certain limit as the normal stress increases, as it is seen near the point of chip-separation, but it cannot be higher than the limit even though the normal stress increases, the limit being considered to be the shear strength of the material cut. It is quite probable that the friction between tool and chip under such a severe condition as in metal machining differs from the ordinary friction where Coulomb's law holds. The chip is pressed so severely against the rake face that welded patches of the material will be spread all over the apparent tool-chip contact area on the rake face. This will result in the aforementioned fact that the frictional stress on the rake face is almost equal to the shear strength of the material cut. Fig. 9 shows the variation of the coefficient of friction on the rake face, which apparently contradicts the commonly used assumption that the coefficient of friction on a rake face is constant. Fig. 10 shows the photomicrograph of a tool-chip contact area on the rake face. The tool-chip contact length measured from the photomicrograph is exactly equal to the distance from the cutting edge to the position of the zero fringe in Fig. 5. One can also find out that about one third of the contact area from the cutting edge is dimmer than the other part. The former part is covered by numberless fine welded patches by which the chip seems to be adhered strongly to the rake face. It is an interesting fact that the width of this part is identical with that of the region where the normal stress is stationary in Fig. 8. The attribute of this region may have something to do with formation of a built-up edge.

Distribution of Normal Stress on Shear Plane

Both the direction and the acting point of the resultant cutting

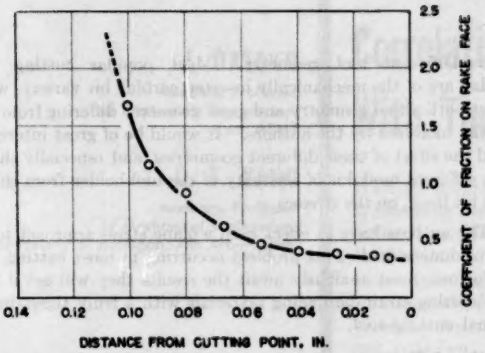


Fig. 9 Distribution of coefficient of friction on rake face. Cutting conditions same as Fig. 4.

force can be figured out by the following two equations taking account of the stress distributions on the rake face which have been determined already

$$S = \frac{\int_A^B \sigma(x)x\,dx}{\int_A^B \sigma(x)\,dx} \tag{5}$$

$$\tan\theta = \frac{\int_A^B \tau(x)\,dx}{\int_A^B \sigma(x)\,dx} \tag{6}$$

where S denotes the distance from the cutting edge to the acting point of the resultant force, $\sigma(x)$ and $\tau(x)$ the normal and frictional stress on the rake face, respectively, and θ the angle between the rake face and the resultant force. The calculation of each definite integral was conducted graphically from the cutting edge A to the point of chip-separation B. θ and S thus calculated are 64 deg and 0.038 in., respectively.

Fig. 11 shows a schematic diagram of orthogonal cutting, in which B is the point of chip-separation, P_1 the acting point of the resultant force, and R_1 the vector of the resultant force. The acting point Q_1 of the resultant force in the shear plane must be located on the acting line of the resultant cutting force R_1 on the rake face, because both forces must be in equilibrium. The distance from the cutting edge to Q_1, which can be easily computed by using the angle relationships of the chip geometry, is 0.048 in.

If the stress distributions on both the shear plane and the rake face are uniform as have been hitherto considered, the two acting points Q_2 and P_2 of the resultant force R_2 must be located at the centers of the shear plane and the tool-chip contact length, respectively, as shown in Fig. 11. The real situation is not so, as has been calculated, but the acting point Q_1 in the shear plane is upwardly shifted from the center Q_2 of the shear plane as shown in Fig. 11. It can be concluded that the distribution of the normal stress in the shear plane is not uniform, but the normal stress increases as a whole toward the work surface, because the position of the acting point of the resultant force in the shear plane has nothing to do with the distribution of the shear stress in the shear plane. In Fig. 11, the supposed distribution of the normal stress σ_{s1} in the shear plane is shown by the broken lines while the uniform model σ_{s2} is illustrated by the dotted lines.

Conclusions

1 The stress distributions on the rake face could be successfully determined photoelastically by using a tool made of epoxy resin.

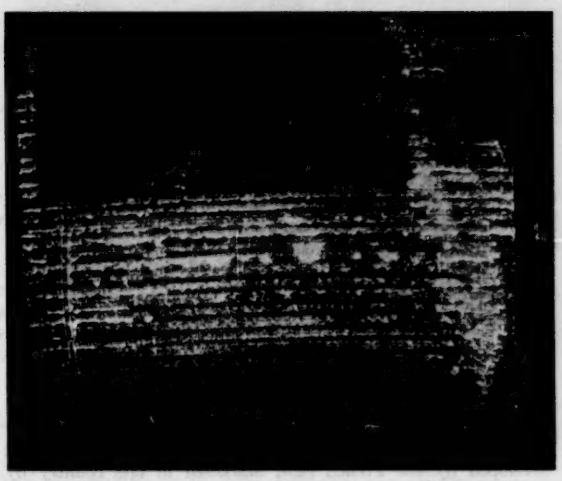


Fig. 10 Photomicrograph of tool-chip contact area. Cutting conditions same as Fig. 4.

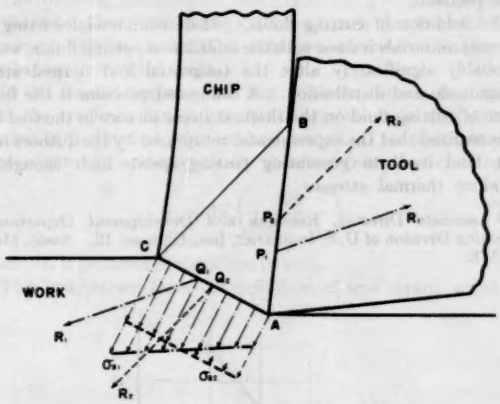


Fig. 11 Diagram of cutting forces and stresses in orthogonal cutting

2 The frictional stress on the rake face is uniform over a wide range of the tool-chip contact area regardless of the varying normal stress thereon, but it decreases steeply near the point of chip-separation.

3 As to the normal stress on the rake face, it has the peak at the cutting edge, in the middle part of the tool-chip contact length it is rather stationary, and in the outermost part it decreases almost linearly vanishing at the point of chip-separation.

4 The peak normal stress at the cutting edge which has been proved by the photoelastic experiment corresponds to the singular force at the cutting edge reported in the authors' previous paper.

5 The distribution of the normal stress in the shear plane is not uniform, but the normal stress increases as a whole toward the work surface.

References

1 H. Takeyama and E. Usui, "The Effect of Tool-Chip Contact Area in Metal Machining," TRANS. ASME, vol. 80, 1958, pp. 1089–1096.

2 E. G. Coker and N. F. Filon, "Treatise on Photoelasticity," 1931, p. 330.

3 M. Okoshi and S. Fukui, "Studies of Cutting Action by Means of Photoelasticity," *Journal of The Society for Precision Mechanics of Japan*, vol. I, no. 15, 1934, p. 598.

DISCUSSION

John K. Liu[3]

I wish to commend the authors on their significant contribution to the understanding of the complex subject of metal cutting. My comments that follow are not meant to be critical and I hope they are not taken as such; they are merely meant to suggest to the authors lines along which they might extend the research in which they have made such an excellent start.

As I see it, there are three significant effects that are not covered in the present investigation. It would certainly be of interest to determine the results of:

1 Cutting strain-hardening materials. Most engineering materials are strain-hardening, and the tool stresses resulting therefrom would be of practical interest. The authors mention the difficulty of obtaining photoelastic material hard enough to cut strain-hardening materials. The birefringent plastic coating developed by the French (and marketed in this country by Tatnall Measuring Systems Co., Phoenixville, Pa.) which lends itself to direct spraying on metals would seem to be a solution to this problem.

2 Addition of cutting fluids. Most commercial cutting on ferrous materials is done with the addition of cutting fluids, which probably significantly alter the tangential and normal stress magnitude and distribution. A connected problem is the influence of cutting fluid on the thermal stress picture in the tool but it is realized that the experimental setup used by the authors may not lend itself to generating cutting speeds high enough to produce thermal stresses.

[3] Associate Director, Research and Development Department, Clearing Division of U. S. Industries, Inc., Chicago, Ill. Assoc. Mem. ASME.

3 Different tool geometry. Most popular cutting tools today are of the mechanically inserted carbide bit variety, which offer both a tool geometry and gross geometry differing from that of the tool used by the authors. It would be of great interest to find the effect of these different geometries, and especially that of the different modulus of elasticity of the tool holder from that of the bit itself, on the stresses.

The authors have in effect used a plane stress approach to the three-dimensional stress problem occurring in most cutting, and I, for one, most anxiously await the results they will get if they try cutting strain-hardening materials with a truly three-dimensional cutting tool.

Authors' Closure

The authors are indebted to Mr. Liu for his helpful suggestions. His proposal to use the birefringent plastic coating offers interesting possibilities. The coating will be applicable to both the side of two-dimensional cutting tool made of tool steel instead of photoelastic material and the side of workpiece. Such an approach is now in progress in the authors' laboratory and it is expected to bring about helpful information for the understanding of the machining stresses.

The authors are also quite interested in inquiring about the effect of cutting fluids on the tool-face friction. Using the same setup described in this paper, it was found by applying a mineral oil that the tool-chip contact length is greatly reduced and the normal stress on the rake face is increased very steeply toward the cutting edge and it reaches a higher value than that in dry cutting. Because most cutting fluids are not so effective for lead, as is well known, the authors are checking the results in cutting of babbitt which was also found to be machinable by the resin tool.

J. DATSKO
Associate Professor
of Mechanical Engineering.
Mem. ASME

C. T. YANG
Associate Professor
of Mechanical Engineering,
University of Michigan, Ann Arbor, Mich.

Correlation of Bendability of Materials With Their Tensile Properties

With the advent of high-temperature-resistant materials in the sheet-metal industry, it is becoming increasingly important to treat the subject of minimum or "critical" bend radius analytically. A simple equation is presented that correlates the minimum bend radius with the percentage reduction of area of the material. The theoretical derivation as well as experimental data are given, with very good agreement between the two. Consequently, it is possible to predict the minimum bend radius for a specific material, provided that the percentage reduction of area, as determined by a standard tensile test, is known. The relationship applies equally well to metals and nonmetals.

Introduction

In the fabrication of parts by plastically bending sheets or bars of a given material, it is necessary to know to what radius R a piece of thickness t may be bent without having fracture occur during the bending process. This has always been done experimentally in sheet-metal shops by using various tables of R/t values for commonly used materials. However, with the advent of high-temperature-resistant materials it is becoming increasingly important to treat the subject analytically.

About five years ago several researchers in the aircraft industries reported very good correlation between the true strain at the time that the maximum load was applied to a tensile specimen and the R/t ratio to which the same material could be bent. One of the more complete reports of this type is by Paulson, Anderson, and Roberts [1].[1] Fig. 1 of the present paper shows one of their experimentally determined curves which indicates good correlation between R/t and ϵ_m'. However, repeated tests conducted in the University's Mechanical Engineering Materials and Processes Laboratory for the past two years could not substantiate these published data and, therefore, that relationship is not believed to be very reliable.

This past year a new theory was developed by the authors and has been substantiated by many laboratory tests. This theory states that failure will occur in the outer fiber of a material being bent when the true strain in the outer fiber is equal to the true strain at the instant of fracture of a tensile test specimen of the same material. Inasmuch as the true strain at fracture can be very easily determined from the conventional percentage reduction of area, the application of this theory is greatly simplified as will be shown.

Analysis

The following assumptions were made: (*a*) The fracture strain in the outer fiber of a bending specimen equals that in tensile test specimen, (*b*) the material is homogeneous and isotropic, (*c*) the bar bends in plain strain.

In the tensile specimen, see Fig. 4, the maximum true strain that the specimen can endure is the true strain in the reduced section at the instant prior to the fracture. This strain may be very accurately and very easily obtained from the relationship

$$\epsilon_f = \ln\left(\frac{100}{100 - A_r}\right) \tag{1}$$

where A_r is percentage reduction in area.

This is apparent from the definition of true strain, which is,

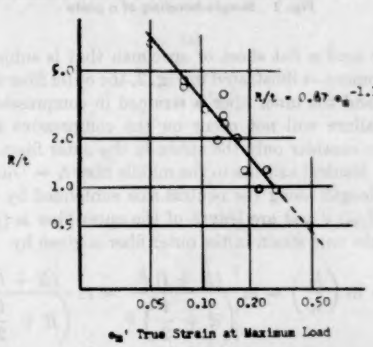


Fig. 1 *R/t* **versus critical strain (true strain) at the maximum load of a tensile test as reported by D. L. Paulson, W. E. Anderson, and E. C. Roberts**

[1] Numbers in brackets designate References at end of paper.

Contributed by the Production Engineering Division and presented at the Annual Meeting, Atlantic City, N. J., November 29–December 4, 1959, of The American Society of Mechanical Engineers.

Note: Statements and opinions advanced in papers are to be understood as individual expressions of their authors and not those of the Society. Manuscript received at ASME Headquarters, August 14, 1959. Paper No. 59—A-110.

Nomenclature

ϵ = true strain
ϵ_m' = true strain at maximum tensile load
ϵ_f = true strain at fracture
ϵ_o = true strain in outer fiber
ϵ_i = true strain in inner fiber
A_r = percentage reduction of area in tensile specimen
l_f = final length
l_o = original length
A_f = final cross-sectional area
A_o = original cross-sectional area
t = thickness of the bent bar
θ = angle subtended by an arc length in a bent bar
R_o = radius at outer fiber of the bent bar
R = radius at inner fiber of the bent bar
R_n = radius at the neutral axis of the bent bar
k = ratio defined as R/t
w = width of specimen
F = force applied
D = diameter of rolls in bending test
d = distance between rolls in bending test
M = bending moment

"the summation of changes in length divided by the length from which each change was produced." In mathematical form this is expressed as

$$\epsilon = \frac{\Delta l_1}{l_o} + \frac{\Delta l_2}{l_o + \Delta l_1} + \frac{\Delta l_3}{l_o + \Delta l_1 + \Delta l_2} + \dots \quad (2)$$

By letting Δl become vanishingly small, the expression for true strain then reduces to

$$\epsilon = \ln \frac{l_f}{l_o} \quad (3)$$

where ln is the natural logarithm, l_f is the final length, and l_o is the original length.

Also, if the volume of the material is assumed to remain constant with only its shape being changed, then the ratio of A_o/A_f must equal the ratio of l_f/l_o and then Equation (3) may be expressed as

$$\epsilon = \ln \left(\frac{wt}{w_f t_f} \right) = \ln (A_o/A_f) \quad (4)$$

Since, by definition, $A_r = (A_o - A_f)/A_o \times 100$, which may be rewritten as $A_o/A_f = 100/(100 - A_r)$. If this latter expression is substituted in Equation (4), and the strain ϵ is defined as the true strain at fracture, ϵ_f, then Equation (1) results.

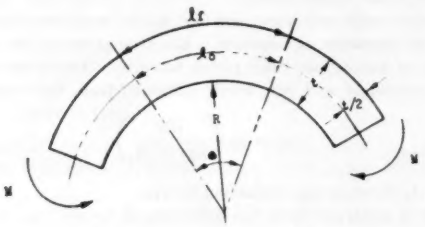


Fig. 2 Simple bending of a plate

Consider next a flat sheet or specimen that is subjected to a bending moment as illustrated in Fig. 2, the outer fiber is strained in tension and the inner fiber is strained in compression. Inasmuch as failure will not occur on the compressive side, it is sufficient to consider only the strain in the outer fiber.

Case 1. Neutral axis lies in the middle fiber $t_i = 1/2t$.

The arc length along the neutral axis subtended by the angle θ is $(R + 1/2t)\,\theta$ and arc length of the outer fiber is $(R + t)\,\theta$. Therefore the true strain in the outer fiber is given by

$$\epsilon_o = \ln \left(\frac{l_f}{l_o} \right) = \ln \frac{(R + t)\,\theta}{\left(R + \frac{t}{2}\right)\theta} = \ln \frac{(R + t)}{\left(R + \frac{t}{2}\right)} \quad (5)$$

This is based on the assumption that the neutral axis does not deviate appreciably from the half-thickness of the sheet, which is very valid for R/t ratios greater than $1^1/_2$. By equating ϵ_o in Equation (5) to ϵ_f in (1), it is apparent that

$$\frac{100}{100 - A_r} = \frac{R + t}{R + \frac{t}{2}} \quad \text{or} \quad \frac{R}{t} = \frac{50}{A_r} - 1 \quad (6)$$

From the authors' experience it was found that this relationship is valid for materials having a percentage reduction in area (A_r) of less than 20. For materials having a percentage reduction in area greater than 20, the displacement of the neutral axis in bending is significant and will be discussed next.

Case 2. Neutral axis is not at the mid layer.

$$\epsilon_o = \ln \left(\frac{R_o}{R_n} \right) \quad (7)$$

If k is defined as R/t, one has

$$\left.\begin{aligned} R_o &= t(k + 1) \\ R &= kt \\ R_n &= t(k^2 + k)^{1/2} \end{aligned}\right\} \quad \dots \text{from Hoffman and Sachs [2].} \quad (8)$$

Substituting Equation (8) into (7) gives

$$\epsilon_o = \ln \frac{k + 1}{(k^2 + k)^{1/2}}$$

or

$$\epsilon_o = \ln \left(1 + \frac{1}{k} \right)^{1/2} \quad (9)$$

From Equation (8)

$$k = \frac{R_o}{t} - 1 \quad (10)$$

Combining Equations (9) and (10), one obtains

$$\frac{R_o}{t} = \frac{1}{e^{2\epsilon_o} - 1} + 1 \quad \text{and} \quad \frac{R}{t} = \frac{1}{e^{2\epsilon_o} - 1} \quad (11)$$

From Equation (1), $e^{\epsilon_o} = 100/(100 - A_r)$, where $\epsilon_o = \epsilon_f$. By substituting e^{ϵ_o} in Equation (11), the following equation results

$$\frac{R}{t} = \frac{(100 - A_r)^2}{200A_r - A_r^2} \quad (12)$$

The plotting of Equations (6) and (12) is shown in Fig. 3.

It is seen that both curves look hyperbolic and are asymptotic to the R/t axis. In general for the same percentage reduction of area the solid curve gives lower value of R/t than the dotted curve and the difference becomes smaller as A_r becomes smaller in value and vice versa. When A_r is less than 4, the two curves practically coincide.

Also, from laboratory experience, it has been found that the empirical relationship

$$\frac{R}{t} = \frac{60}{A_r} - 1 \quad (13)$$

is a very satisfactory relationship for the ductile materials that do have a shift in the position of the neutral axis.

Experimental Technique and Results

Flat tensile specimens were cut from magnesium, aluminum, brass, 1018 steel, titanium, cast-iron, and plastics bar stock. In nearly all cases, three specimens were tested and since the range of values obtained for each material was not large, only the averages are reported. The composition and code of each are shown in the third column of Table 1, and the dimensions of the flat tensile specimens are shown in Fig. 4. To determine the agreement between the tensile properties obtained from flat and round tensile specimens of the same material, round tensile specimens for some metals were also tested, and the percentage reduction of area is also shown in Table 1.

The corresponding bending specimens were cut from the same bar stock to the dimensions shown in Fig. 5.

A 30-ton universal tensile-testing machine was used to pull the tensile specimens until they fractured. The cross-sectional areas of the specimens before testing and after fracture were measured on a toolmaker's microscope with an accuracy of 0.0001

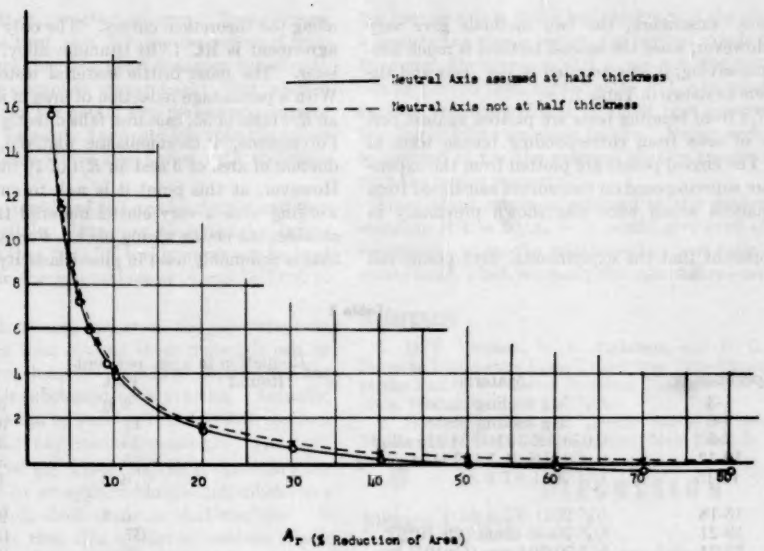


Fig. 3 Curves from analytical solution for R/t versus percentage reduction of area

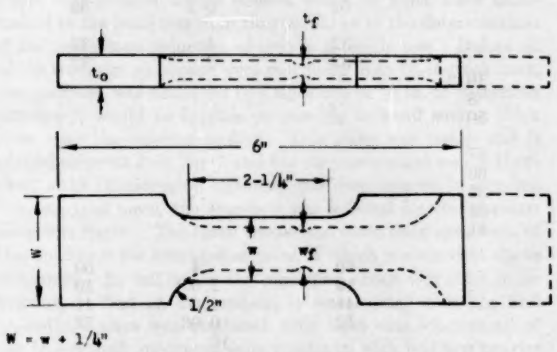


Fig. 4 Dimensions of flat tensile testing specimens

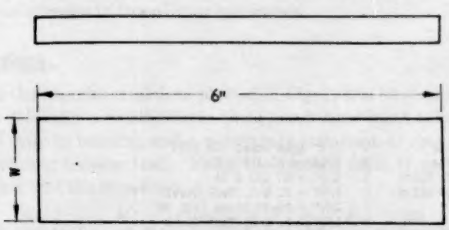


Fig. 5 Dimensions of a bending specimen

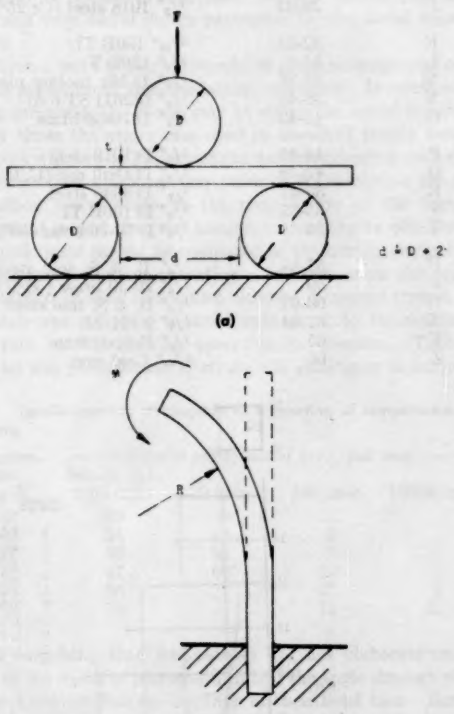


Fig. 6 Bending fixtures used for testing

in. The percentage reduction of area for each specimen was calculated and the results are listed in column 7 of Table 1.

In the bending test two methods were tried. In the first method the specimen was simply supported on two rolls and the load was applied through a third roll to deflect the flat specimen as shown in Fig. 6(*a*). The set of bottom rolls was selected such that $d = D + 2t$ [refer to Fig. 6(*a*)]. Ten sets of such rolls ranging from $^1/_8$ in. diam to $1^1/_4$ in. diam were prepared for the test. During the test it was necessary to stop loading and search for the small cracks on the outer fiber of the beam. Whenever the cracks appeared, the test was stopped and the inner radius of the bent beam was measured by means of a set of radius gages.

In the second, more expedient, method of bending the flat bar was clamped in a vise vertically on one end as a cantilever beam and a bending moment was applied manually on the other end. The moment should be applied carefully so that the beam will be subjected to bending only and no shear forces exerted onto the bar. Fig. 6(*b*) shows this method schematically. During the process of bending, careful observation is also required to see the initiation of small cracks. Once the small crack appears, the loading should be stopped and the inner radius of the specimen measured as in the first method.

From the authors' experience, the two methods give very similar results. However, since the second method is much simpler to use and time-saving, it was adopted in the tests and the results obtained were as shown in Table 1.

The results of R/t from bending tests are plotted against percentage reduction of area from corresponding tensile tests as shown in Fig. 7. The circled points are plotted from the experimental data and are superimposed on two curves calculated from the theoretical analysis which were also shown previously in Fig. 3.

It is at once apparent that the experimental data points fall along the theoretical curves. The only material not giving close agreement is RC 130B titanium alloy. This will be explained later. The most brittle material tested was a gray cast iron. With a percentage reduction of area of slightly less than one and an R/t ratio of 55, cast iron falls directly on the theoretical curve. Polystyrene, a thermoplastic material, with a percentage reduction of area of 3 and an R/t of 16 fits the curve equally well. However, at this point it is well to emphasize the need, when working with a very elastic material that has no plasticity, to consider the elastic spring back. For example, CR39 is a plastic that is commonly used in photoelasticity studies and as normally

Table 1

	Specimen no.	Material	Reduction in area, per cent Round	Reduction in area, per cent Flat	$\frac{R}{t}$	Reduction in area, per cent
A	1-3	1/4" Mg tooling plate	...	9.9	3.5	9.9
B	4-6	1/8" Mg tooling plate	...	11	3	11
C	7-9	0.050 AZ 31 BH 24 Mg alloy	...	19	3.7	19
D	10-12	0.032 HK 31 X1 H24 Mg alloy	...	25	3	25
E	13-15	1/4" 2011 ST 6 Al	...	6.2	10	6.2
F	16-18	1/8" 2011 ST 6 Al	...	6.4	9	6.4
G	19-21	1/4" 70-30 brass (Cr 10%)	...	37	0.5	37
H	22-24	1/8" 70-30 brass (Cr 10%)	...	40	0.55	40
I	26-28	1/4" 1018 steel (Cr 25%)	...	53	0.2	53
J	29-31	1/8" 1018 steel (Cr 25%)	...	52	0.2	52
K	32-33	3/16" 130B Ti	...	32	4.	32
L	34	1/4" 130B Ti	...	32	2	32
M	35-37	1/8" D Mg tooling plate	10.2	...	...	...
N	38-40	1/8" D 2011 ST 6 Al	8	...	...	...
O	41-43	1/8" D 70-30 brass	50	...	...	...
P	44-46	1/8" D 1018 steel	57	...	...	...
Q	47-49	1/8" D Drill rod (C.T.S.)	60	...	...	...
R	50-52	1/8" D 4340 steel	35	...	...	...
S	53-55	1/8" D 130B Ti	39	...	...	...
T	56	1/8" D 70-30 brass (50% CW)	41	...	...	...
U	57-58	1/8" B. & S. tool steel	...	60	.4	60
V	59	1/4" 70-30 brass (50% CW)	...	19	2.0	19
W	60-61	1/4" B. & S. tool steel	...	41	1.0	41
X	62-64	1/4" 2 SH Al	...	52	0.06	52
Y	65	1/8" Polystyrene	...	...	16	3
Z	66	9/64" Cast iron	...	...	55	3/4

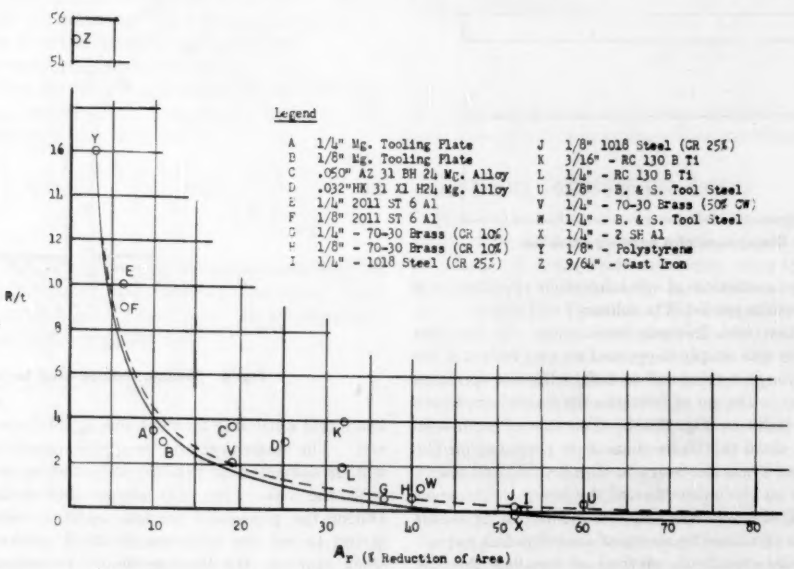

Legend
A 1/4" Mg. Tooling Plate
B 1/8" Mg. Tooling Plate
C .050" AZ 31 BH 24 Mg. Alloy
D .032"HK 31 X1 H24 Mg. Alloy
E 1/4" 2011 ST 6 Al
F 1/8" 2011 ST 6 Al
G 1/4" - 70-30 Brass (CR 10%)
H 1/8" - 70-30 Brass (CR 10%)
I 1/4" - 1018 Steel (CR 25%)
J 1/8" 1018 Steel (CR 25%)
K 3/16" - RC 130 B Ti
L 1/4" - RC 130 B Ti
U 1/8" - B. & S. Tool Steel
V 1/4" - 70-30 Brass (50% CW)
W 1/4" - B. & S. Tool Steel
X 1/4" - 2 SH Al
Y 1/8" - Polystyrene
Z 9/64" - Cast Iron
R/t
A_r (% Reduction of Area)

Fig. 7 Experimental data superimposed on analytical curves of R/t versus percentage reduction of area

tested has a zero percentage reduction of area. Nevertheless, a $^1/_8$-in-thick sheet can be bent to a 2-in. radius before it fractures. However, after breaking, the bend specimen is perfectly flat or straight indicating complete spring back. But when a tensile specimen is tested and the actual area measured just prior to fracture, it is found to be nearly 3 per cent less than the initial area, and with this correction CR39 also agrees with the theoretical curves.

Normally, aluminum is considered to be quite ductile and thoriated magnesium quite brittle. However, by comparing points E-F to C-D it is apparent that 2011 ST 6 aluminum can be bent to an R/t of 9 or 10 while the magnesium alloy can be bent to an R/t of only 3 to 4.

For the soft brasses, aluminum, and steel, the percentage reduction of area is greater than 50 and these materials can be bent to an R/t ratio approaching 0. This is predicted from the theoretical equations and is substantiated in practice. Actually, these soft ductile materials can be bent over double with one face in contact with itself without any cracks appearing.

The experimental data of the $^3/_{16}$-in. RC 130B titanium alloy is off the theoretical curve by an appreciable amount, much more so than the one piece of $^1/_4$-in-thick titanium that was bent. It is believed by the authors that this is due to residual tensile stresses set up during the machining of the $^3/_{16}$-in-thick specimens. These very shallow tensile stresses would be much more detrimental in the bend test than they would be in the determination of the percentage reduction of area in a tensile test. Before all of the titanium specimens were machined from $^1/_2$-in-thick bars, one specimen was machined to a thickness of $^1/_4$ in. to determine whether it would be feasible or possible to bend such a thick piece with the existing tooling. This piece was tested and is plotted as point L in Fig. 7 and fits the curve quite well. However, since considerable difficulty was encountered in bending this one trial piece, the standard size selected for the titanium alloy was $^3/_{16}$ in. The three tensile and three bend specimens of this thickness are averaged as point K which is somewhat above the curve. In talking to the machinist about this later in an attempt to find an explanation, it was learned that the first $^1/_4$-in-thick piece was machined with light cuts whereas all of the $^3/_{16}$-in-thick specimens were machined with just one heavier cut taken from each side, which could account for the presence of residual stresses in these latter specimens.

Conclusions

From the experimental data plotted in Fig. 7, it is obvious that there is definitely a correlation in theory and experiment between the R/t ratio in bending and a percentage reduction of area in a corresponding tension test. From the plotted data, it may be concluded that the equations

$$\frac{R}{t} = \frac{50}{A_r} - 1, \quad \frac{R}{t} = \frac{(100 - A_r)^2}{200\,A_r - A_r^2}, \quad \text{or} \quad \frac{R}{t} = \frac{60}{A_r} - 1$$

can be used for the prediction of a minimum bend radius with respect to a particular plate thickness. Usually this property of a material is given in a handbook. Even if the data are not available in publications, it is quite simple to obtain from a tensile test.

In the derived equation of $R/t = 50/A_r - 1$ the only property which affects the bend radius of a material is A_r and nothing else. Therefore this equation is valid for all materials, metal or nonmetal. This is confirmed by the experimental data of two plastics which fall directly onto the theoretical curve. However, the authors have some reservation for generalizing this theory because of the limited data for nonmetals tested.

It is understood that ϵ_o (strain in outer fiber) and ϵ_i (strain in inner fiber) during bending are equal in magnitude only when the neutral axis is at the half thickness of the plate. If the beam is bent in the plastic region, the neutral axis is no longer at mid-fiber and, therefore, ϵ_o and ϵ_i are not the same. However, it can be proved that, for the same R/t, ϵ_o is much higher than ϵ_i. In fact, from practice it is found that compressive regions in bending hardly ever cause failure. Under such circumstances it is justified to use the extreme fiber in tension side to correlate with the tensile property.

Since plane strain is assumed in the analysis, the proposed equation $R/t = 50/A_r - 1$, would give even closer results with experiment when the width-to-thickness ratio of the plate becomes large, which is usually the case in sheet-metal work.

References

1 D. L. Paulson, W. E. Anderson, and E. C. Roberts, "Metal Forming Parameters From True-Stress True-Strain Data," presented at the Pacific Northwest Regional Conference of the AIMME, Spokane, Wash., April 28–30, 1955.

2 Oscar Hoffman and George Sachs, "Theory of Plasticity," McGraw-Hill Book Company, Inc., New York, N. Y., 1953, p. 268.

DISCUSSION

Richard L. Kegg[2]

The authors have done an excellent job of investigating the bendability of materials and their findings should provide a unique and very useful design parameter for the metal working industry.

However, a word of caution should be given to design engineers who will make use of this bendability criterion. In some metal forming processes, the strain rate at which the metal is formed is many times the strain rate used in standard tensile testing. This point will not be important for most engineering materials, but for materials which are very strain-rate sensitive or are near a transition temperature at the temperature of the forming operation, it means that the maximum ductility as obtained in slow tensile tests cannot be compared to the maximum ductility in a high-speed forming operation. In these cases the tensile test reduction of area at fracture must be measured from a test bar which was pulled at a strain rate equal to the maximum strain rate in the bending operation in question. Table 2[3] illustrates this phenomenon of strain rate sensitivity in tungsten.

Table 2 Tensile ductility of tungsten as a function of temperature and strain rate

Temperature, deg F	Tensile reduction of area, per cent — Strain rate 0.05/min	10/min	100/min	19000/min
302	20	0	..	..
392	34	42	0	..
482	40	53	1	..
572	47	60	52	0
707	56	..	67	5
932	70	..	71	62

It is surprising that the authors did not elaborate on the effects of the width of test specimen and the angle through which it is bent on the limiting ductility in their bend test. Barlow[4] has reported that these quantities have a definite effect on deformation in bending. According to his findings, the ductility should be less for wider workpieces and end effects should be greater for smaller angles of bend.

In any event, these factors do not seem to have had much effect on the ultimate results of the bend testing in this particular

[2] Senior Research Engineer, The Cincinnati Milling Machine Company, Cincinnati, Ohio. Assoc. Mem. ASME.

[3] A. W. Magnusson and W. M. Baldwin, *Journal of Mechanics and Physics of Solids*, vol. 5, 1957, pp. 172–181.

[4] D. A. Barlow, "Formability of Aluminum Alloys," *Engineering*, vol. 181, 1956, p. 366.

application as evidenced by the good agreement which the authors found between experimental data and simple theory. The authors are to be congratulated for making a significant contribution to the knowledge of the metal working industry.

Authors' Closure

The authors wish to thank Mr. Kegg for his kind discussion and compliments. They agree wholeheartedly with Mr. Kegg that for strain-rate sensitive materials the strain-rate effect should be carefully incorporated in the bendability criterion. In fact, this is the second step of the study the authors wished to pursue.

In the tests the authors limited their work to thin and narrow sheets. For thick and wide plates the criterion may not hold. This is another phase of the problem which should be carefully investigated in the future. It is hoped that this paper will stimulate others to study some of these related problems.

Comparisons Between the Shearing Properties of Alpha-Brass as Derived From the Cutting Process and From Static and Impact Torsion Tests[1]

S. S. CHANG[2]
W. B. HEGINBOTHAM[3]

Introduction

THE METAL-CUTTING PROCESS has been recognized as a shearing process for some considerable time. Nevertheless, the conditions under which shearing takes place and the behavior of the work material in the region of shear is still not fully understood.

It is certain that such factors as temperature and strain rate will affect the shearing characteristics of a material and both these factors are present in the metal-cutting operation. The presence of a compressive stress applied normal to the plane of shear is also thought to have some effect. This can perhaps be visualized as due to the presence of an "internal frictional coefficient" within the structure of the material itself [1].[4] On the other hand Shaw [2] holds the opinion that the size and extent of the shearing zone can have some influence. Both these points of view furnish possibilities for the explanation of the lack of correspondence between shearing flow stresses obtained by analyzing the cutting process [3], and those obtained from conventional torsion tests. There is also the question of the effects of straining at very high rates, but no published information to date suggests that there is any large increase in the *ultimate* flow strength of alpha-brass under such conditions.

It is thus difficult to explain why stresses derived from cutting tests on α-brass are from 40 to 80 per cent higher than the characteristics of the material would suggest, and the question which springs to mind is this: Are these differences real or is there something amiss with our methods of analysis?

How the variables, temperature and strain rate (which are somewhat dependent on each other under certain conditions), together affect the characteristics of the cut material is largely a matter for speculation. Shaw and Finnie [2] put forward the hypothesis that the combined effects of strain rate and temperature cancel, whereas Chao and Bisacre [4] and Chao, Trigger, and Zylstra [5] assert that, because the time taken for material to traverse the shearing zone is so short and consequently the time for heating is so small, the properties of the material should be appropriate to those found at normal room temperatures.

Until recently, direct comparisons between the shearing properties of a material derived from cutting tests and those derived from tests on the same material, where certain features of the shearing conditions of the cutting process are imitated, have not been available. Some work with this end in view has been carried out by Bastien and Weisz [6] using steel as a work material.

In the work presented here direct comparisons were made between the results obtained by analyzing the cutting process and the results of both static and impact torsion tests on the same material. The material used was alpha-brass (70 to 30 per cent Cu-Zn). Choice rested on this material because it appeared to exhibit a very large internal frictional coefficient if results obtained from an "orthogonal" cutting test were analyzed in terms of the now familiar Merchant analysis. These cutting tests indicated that the sum $2\phi + \tau - \alpha = 67$ deg $\pm$ 5 per cent if the correct cutting conditions were chosen. The difference between 67 deg and the ideal figure of 90 deg, which should be the case if the material exhibits no Bridgman effect, should therefore be easy to find. It is shown that this difference could not be accounted for by the internal frictional coefficient effect.

Due to the failure to explain the difference, efforts were shifted toward an evaluation of the possible effects of rate of strain on the stress-strain curve for the material. Towards this end a special impact-torsion machine was designed capable of impressing strain rates of 1000 per second abruptly to a tubular specimen. This figure for strain rates compares favorably with the lower limit of the estimates of strain rates in metal cutting by another investigator [7]. Photomicrographs covering the range of conditions tested were produced by an abrupt stopping technique and from these it was possible to explain certain variations, particularly in the constant $2\phi + \tau - \alpha = C$. It was not found possible to make any estimate of strain rates from the photomicrographs.

Cutting Tests

Orthogonal cutting tests were carried out in a heavy-duty center lathe on tubular specimens. The alpha-brass material for these tests was supplied in the form of $1^1/_4$-in. square bars which were hot bent into semicircular rings and these were accurately butt-jointed together to form complete rings of $12^3/_4$-in. diameter. High-speed steel tools were used and cutting forces were recorded by a two-component dynamometer in which the deflections of diaphragms supporting the tool cradle were measured by inductance heads and a suitable bridge circuit. Calibration was by means of dead loading with a dummy tool in place and it was found that this could be relied upon to an accuracy of $\pm$4 lb including reading errors. There was negligible "cross sensitivity" of force components applied in the two directions (about 1 per cent of the full load deflection figure). High-speed steel tools were used as these allowed a greater range of rake angle to be explored, and also with brass material high-cutting speeds were

[1] This work was carried out in Manchester College of Science and Technology.

[2] Principal Scientific Officer, National Engineering Laboratory, Glasgow, Scotland.

[3] Senior Lecturer in Production Engineering, University of Nottingham, England.

[4] Numbers in brackets designate References at end of paper.

Contributed by the Metal Processing Activity of the Production Engineering Division and presented at the Annual Meeting, Atlantic City, N. J., November 29–December 4, 1959, of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS.

NOTE: Statements and opinions advanced in papers are to be understood as individual expressions of their authors and not those of the Society. Manuscript received at ASME Headquarters, April 21, 1959. Paper No. 59—A-38.

possible even with such tools. Chip thickness ratios were computed from the assumptions of constant volume and no end flow by direct measurement of the length of cuttings produced from each of the semi-annuli, and it was noted that there was no significant-force discontinuity as the tool traversed the butt joints in the test rings. Tabulated information from the cutting tests is given in Table 1.

The orthogonal cutting tests produced cutting conditions which appeared to give good approximation to the type 2 chip in the Ernst classification [8]. In order to check that the actual chip formation was indeed applicable to the type of analysis required for the derivation of stress values, several photomicrographs covering the extremes of the speed range chosen were produced as shown on Figs. 1(*a*, *b*, *c*, and *d*). These photomicrographs were produced by an abrupt stopping technique [9] whereby the workpiece was brought to rest with respect to the tool in a period of time of between 0.5 and 0.7 msec. Such decelerations are possible with this device up to a cutting speed of about 500 ft/min.

These photomicrographs show what appears to be a tendency towards segmentation but this is really an exaggeration because it must be remembered that they only represent *one* particular section of a cut which was 0.20 in. wide. In other words, the irregularities on the chip upper surface are not continuous when considered in a direction at right angles to the plane of flow but were made up of a large number of small "hillocks" distributed at random across the chip upper surface due to the influence of grain boundaries on the flow characteristics of the material. This situation is illustrated in Fig. 2. Thus taken as a whole, the type of chip was reasonably continuous (no chip in practice is truly continuous) and capable of representation by the assumption of complete continuity.

Considering the photomicrographs as shown in Figs. 1(*a* and *b*), these are at slow cutting speeds and exhibit features which do not make the single shear plane type of solution applicable. A definite zone can be observed which is narrowest at the tool tip and fans out as it approaches the intersection between the chip upper surface and the advancing workpiece. With the information at hand, it is not possible to analyze the stress situation within such fields. The strain situation will not be at all uniform in such a zone and is completely obscure. Thus there is no chance whatsoever of extracting information about the stress-strain curve with any simple analysis at such slow cutting speeds.

On the other hand, the higher speed photomicrographs [Figs. 1(*c* and *d*)] show transition conditions between the undeformed and the deformed state which appear to give a nearer approximation to a single thin zone and are therefore more in accordance with the single shear plane concept. Unfortunately, it is not possible to make even the shrewdest guess at the width of the "shearing zone" from these records and no attempt could be made to estimate the average shear-strain rates involved.

Measuring the *apparent* values of ϕ, τ, and α over the whole range of the cutting tests from 23 to 1770 ft/min and finding the sum of $2\phi + \tau - \alpha$ made possible the plotting of the graph in Fig. 3. From this it can be seen that C varies systematically, increasing with increase of speed until a constant value is attained. It is thought here that this apparent variation is the result of the change taking place in the mode of chip formation, i.e., the change from the wide deformation zone at the slower speeds to the narrower shearing zone at the higher cutting speeds. It can be seen from Fig. 3 that C is indeed a good constant for cutting speeds above 500 ft/min and the photomicrographs in Fig. 1 show good "single shear plane characteristics" at about this speed. The value for C above the critical speed is 67 deg ± 5 per cent.

Static Compression-Torsion Tests

A combined torsion-internal pressure machine designed by Gill [10] was modified to provide torsion and axial compression. Details of the machine are as shown in Fig. 4 and the tests carried out closely followed those of P. W. Bridgman [1] as follows:

Table 1 Cutting test results

Cutting Speed v(ft/min)	Feed t(in.)	Top Rake α (deg.)	Merchant Constant C (deg.)	Shear Stress S_s(lb/in^2)	Compressive Stress S_c(lb/in^2)	Shear Strain γ
140	0.00303	30	62.2	61 600	44 900	1.9804
280	0.00303	30	64.2	62 550	48 200	1.9100
621	0.00303	30	66.4	65 260	50 600	1.8220
1271	0.00303	30	68.8	64 700	53 300	1.7640
140	0.00303	15	56.1	58 600	45 150	3.0113
280	0.00303	15	59.0	61 900	49 900	2.8089
1271	0.00303	15	67.0	69 500	64 100	2.3870
1706	0.00303	15	68.7	68 000	64 800	2.3290
140	0.00303	0	57.1	58 800	57 300	4.6290
280	0.00303	0	61.6	65 600	70 000	3.8169
1271	0.00303	0	67.8	72 900	78 000	3.0243
1706	0.00303	0	68.8	71 000	77 100	2.9220
23	0.00498	30	63.3	54 800	41 000	1.8511
145	0.00498	30	66.7	57 900	43 300	1.7591
291	0.00498	30	67.7	58 600	43 800	1.7461
616	0.00498	30	68.3	59 500	45 500	1.7061
996	0.00498	30	68.9	59 800	44 700	1.7083
1320	0.00498	30	68.8	61 500	46 200	1.6704
1770	0.00498	30	69.4	58 500	45 000	1.6704
22	0.00498	15	47.1	70 000	38 900	3.2583
57	0.00498	15	51.2	68 900	37 700	2.8566
135	0.00498	15	54.8	68 700	42 500	2.6383
269	0.00498	15	60.0	70 800	48 500	2.4177
924	0.00498	15	66.3	71 500	55 200	2.2221
1223	0.00498	15	65.2	69 500	51 500	2.1968

(a) $V = 25$ ft/min; $\alpha = 15$ deg; $t = 0.005$ in.; X 100

(c) $V = 490$ ft/min; $\alpha = 30$ deg; $t = 0.010$ in.; X 50

(b) $V = 25$ ft/min; $\alpha = 30$ deg; $t = 0.005$ in.; X 100

(d) $V = 425$ ft/min; $\alpha = 30$ deg; $t = 0.012$ in.; X 50

Fig. 1 Photomicrographs of chips from abrupt stopping tests

The test pieces used were made as shown in Fig. 5. From this diagram it can be seen that the element under test was in the form of an annulus 0.020 in. long by 0.020 in. thick, having an outside diameter of 0.65 in. and situated at the bottom of a notch in the tubular specimen. The reasons for this form of specimen was that it afforded test conditions appropriate to an element where stress gradients could be ignored and suitable support (by a mild-steel mandrel) could be arranged to prevent buckling under the action of the axial loading. It is also not possible to twist a thin long tube to any appreciable value of strain without "wrinkling" of the tube due to the compressive stresses induced at 45 deg to the axis of the tube. Care was also taken to insure that the turning specimens and the torsion pieces were in the same physical state by the application of a suitable heat-treatment developed by checking and comparing characteristics both from microsections and small scale tensile tests.

The measurements taken during these tests were torque, angle of twist over the element (this was measured over the notch plus some length of the tube but elastic deflections of the tube could be ignored), end loading, and longitudinal notch contraction by means of the microscopes.

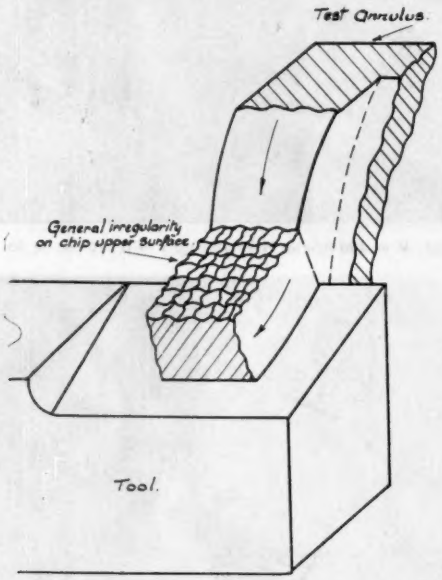


Fig. 2

Fig. 6 shows a family of torque-twist curves obtained from this rig; each member of the family was appropriate to a particular value of the superimposed axial load. The impression given by these curves is that a strong "internal frictional effect" exists.

However, before this could be correctly interpreted, several corrections covering dimensional changes were necessary as follows:

(*a*) A correction for the original dimensions of each specimen (i.e., deviations from the nominal size due to manufacturing technique).

(*b*) Continuous corrections for the changing section of the notch as the test progressed. As each specimen was twisted it was found that the notch length decreased; this was responsible for two effects. Firstly, the shearing strains produced were not directly proportional to the angle of twist alone. Secondly, the axial flow of material was accompanied by a proportional amount of radial flow away from the internal supporting mandrel. Thus the diameter of the test annulus increased and this increase could be accurately calculated by the assumption that the volume of material between the notch limits was a constant. It was possible therefore to modify the nominal stress values based on the original *element* sizes to actual stress values at any stage of the test.

(*c*) A correction for the "spread" of plastic strain into the thicker parts of the test piece. This plastic flow spreading into the main body of the test piece was due to the phenomenon of strain-hardening, i.e., the torque required to promote further straining of the test element increased with strain after the material had started to flow beyond the limits of its elastic behavior. Therefore certain larger sectioned elements outside the notch limits eventually experienced stresses large enough to produce plastic deformation. The elastoplastic boundary spread along a front which was very nearly radial and so a method of correction based on successive approximations was devised as follows: The affected area was split up into slices as shown in Fig. 7. A stress-strain curve was derived using the true shearing stresses calculated from the actual tube section at any value of torque, and computing strain values on the assumption that there was no spread and using the measured notch widths. Thus the strain values in this case would be over-estimated. However, by using this curve in conjunction with the mean shearing stresses derived for each of the slices in Fig. 7, an increment of angle of twist could be estimated. By making a summation of all these small angles

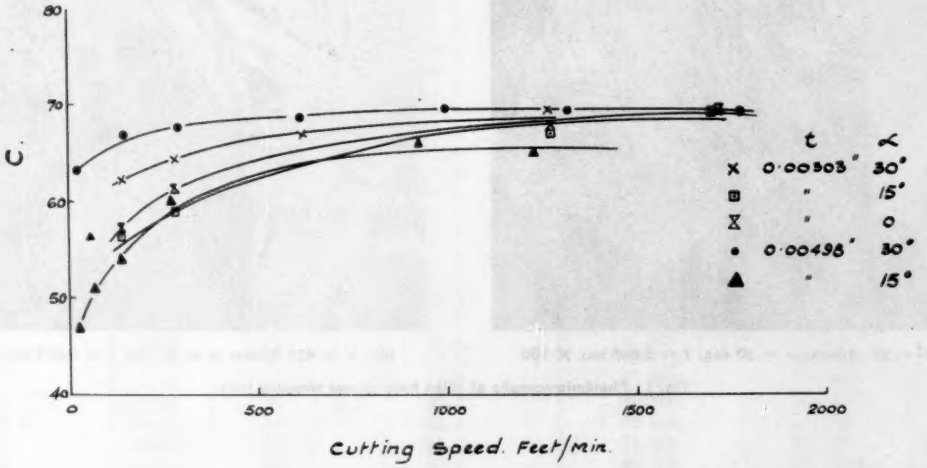


Fig. 3 Merchant constant versus cutting speed

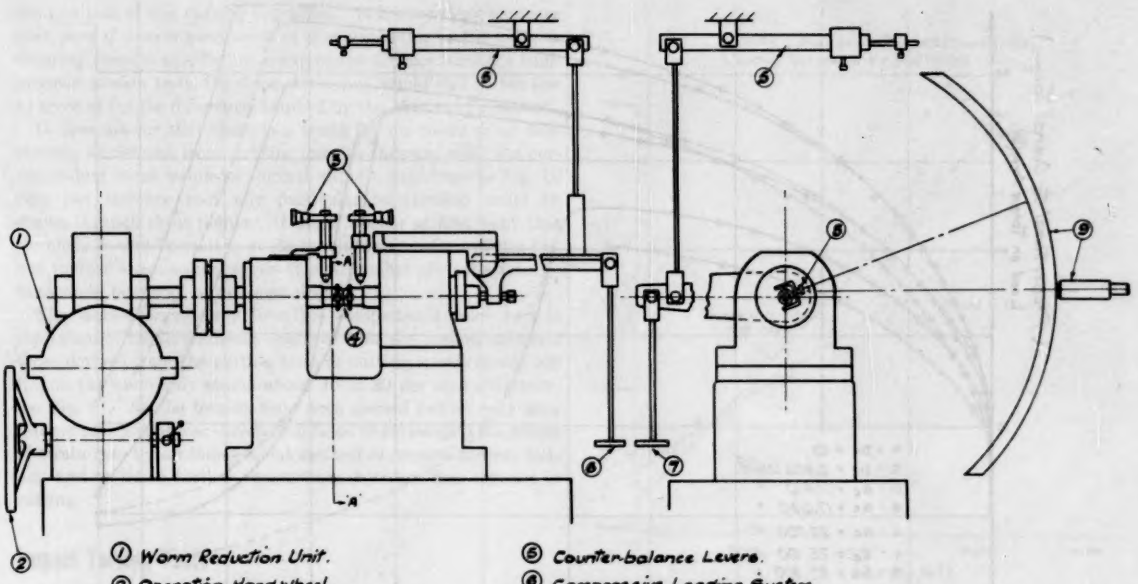

① Worm Reduction Unit.
② Operation Hand Wheel.
③ Microscopes for measurement of Longitudinal Contraction.
④ Specimen.
⑤ Counter-balance Levers.
⑥ Compressive Loading System.
⑦ Torsional Loading System.
⑧ Mirrors.
⑨ Telescopes and Scale for measurement of Twist.

Fig. 4 Static compression-torsion test machine

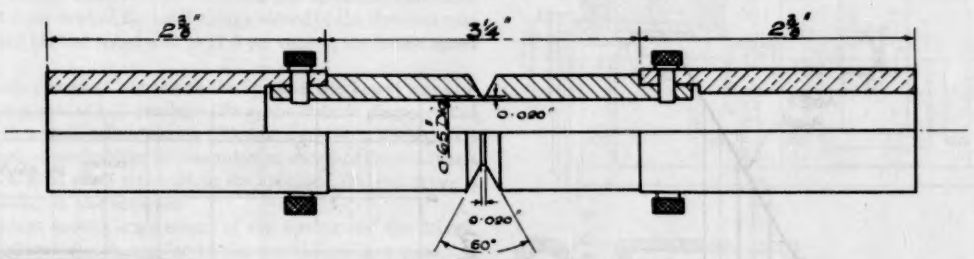


Fig. 5 Specimen for static compression-torsion tests

on each side of the notch a first approximation (over-estimation) to the extra twist involved could be derived. By subtracting this amount from the measured twist values a nearer approximation to the true strain values could be calculated. By applying this new curve a nearer estimation of the actual extra twist could be calculated (slight underestimation). By repeating this process the desired accuracy could be ach'eved.

A set of true shearing stress-strain curves are therefore shown on Fig. 8 and the amount of "spread" correction is indicated. It can be seen that the application of the corrections (*a*), (*b*), and (*c*) just mentioned alter the relative positions of the curves of torque-twist as shown in Fig. 6.

Comparison Between the Cutting Tests and the Static Compression-Torsion Tests

As previously stated, the Merchant constant C in the cutting test at cutting speeds above the critical value of 500 ft/min was found to be 67 deg ± 5 per cent. This suggests a high value for the "internal coefficient of friction" from the static compression-torsion tests. Referring now to the torsion test, Fig. 8 shows that increasing the normal component of stress on the plane of shear serves only to extend the stress-strain curve as a *single function* and not to provide a family of such functions displaced with respect to each other in the direction of the ordinate by proportional values. This agrees with some of the results obtained by Bridgman [1], in particular 1045 steel and this graph is reproduced in Fig. 9.

The difficulty of applying these data to the cutting operation now arises. In order to find a relationship between the shearing stress and the normal stress a particular value of *strain* must be chosen. To compare with cutting tests this could be chosen to represent an average value for the cutting operation. However, applying this to the curve in Fig. 8 no relationship between normal stress and shearing stress would exist as cross plotting would result in one single point only.

Stress-strain points derived from the cutting operation on the basis of a single shear plane concept and a continuous medium are also shown plotted in Fig. 8 and it is seen that the mean shear stresses in cutting are very much above those from the torsion tests (from 40 to 80 per cent).

There *is* an increase in the *ultimate* shearing stress with increasing normal stresses indicated by the compression-torsion tests on Fig. 8. The relationship between ultimate shearing stress and normal stress, irrespective of the strain values at which the ultimate takes place, can be seen in Fig. 10 and this is very

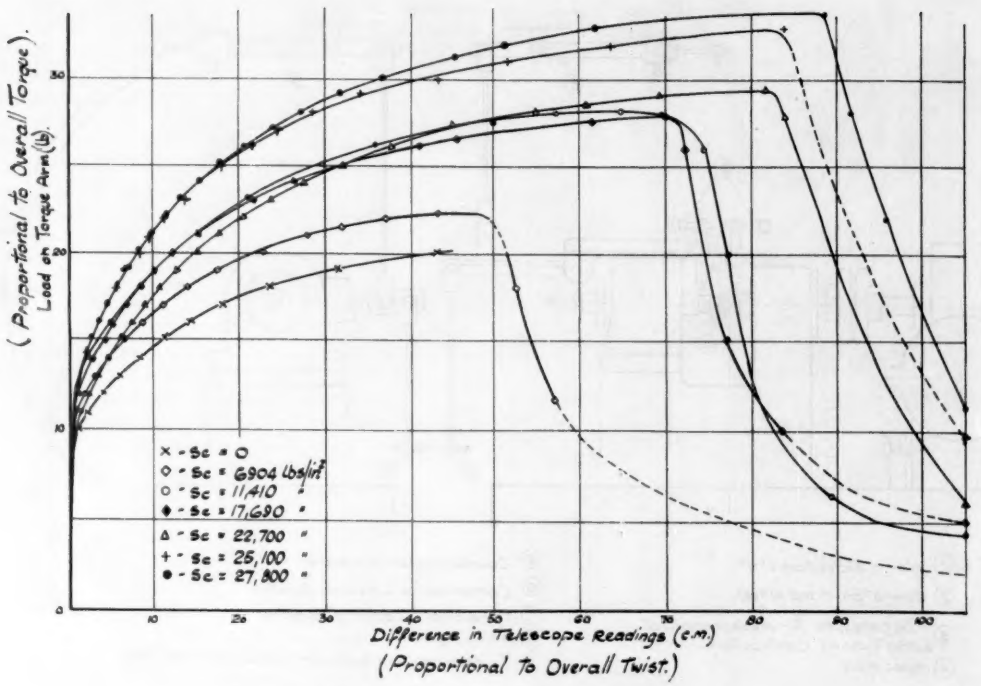

(Proportional to Overall Torque).
Load on Torque Arm (lb).
× - Sc = 0
◇ - Sc = 6904 lbs/in²
○ - Sc = 11,410 "
♦ - Sc = 17,690 "
Δ - Sc = 22,700 "
+ - Sc = 25,100 "
● - Sc = 27,800 "
Difference in Telescope Readings (c.m.)
(Proportional to Overall Twist.)

Fig. 6 Torque-twist curves from static compression-torsion tests

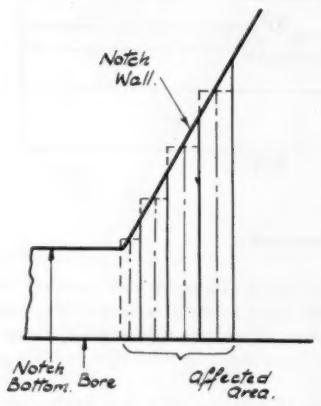


Fig. 7

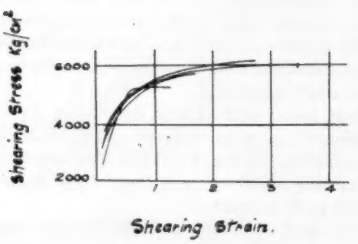


Fig. 9 Shearing stress against shearing strain (conventional) for a 1045 steel. The four curves are for compressive loads of 0, 2300, 4400, and 5700 kg/cm², respectively, in the order of length of the curves. Bridgman [1].

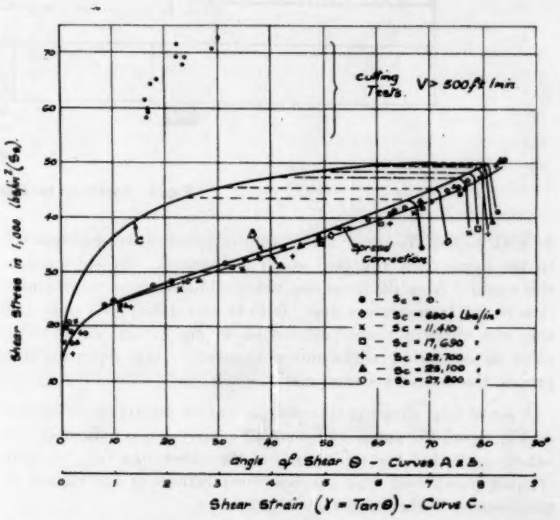

Cutting Tests. V> 500 ft/min
Torsional Correction
Shear Stress in 1,000 lbs/in² (Ss).
Angle of Shear Θ - Curves A & B.
Shear Strain (γ = Tan Θ) - Curve C.

Fig. 8 Comparison of results from cutting and static compression-torsion tests

nearly linear showing a slope of 0.273. Cutting-test results for cutting speeds higher than 500 ft/min are shown plotted on the same graph and, by extrapolation, the compression-torsion relationship is seen to indicate stresses which are from 5 to 20 per cent lower. Putting a line of slope cot 67 deg (0.4245) through the cutting results shows quite clearly the difference between the slope of the ultimate shearing-stress curve and that implied by

the analysis of the cutting operation. It is very clear therefore that, *even if it were permissible* to plot variations in the *ultimate* shearing stresses at different compressive stresses from the compression-torsion tests, the slope of the line would still be too low to account for the differences implied by the Merchant constant.

It does appear that there is a trend for the mean shear flow stresses as derived from cutting tests to increase with the corresponding mean values of normal stress. Reference to Fig. 10 does not indicate that any particular relationship could be drawn through these points. It might appear at first sight that the chain dotted line put in at the required slope of cot 67 deg fits, but further examination shows that a number of "satisfactory" lines could be drawn at different slopes.

The largest discrepancy from the comparisons made here is the values of the flow stresses observed from the torsion tests and those derived from the cutting tests at cutting speeds above 500 ft/min (as previously stated about 40 to 80 per cent difference, see Fig. 8). As the torsion tests were carried out at very slow rates of straining it was therefore decided to investigate the effects of strain rate by an independent method of impact torsion testing, and to make further comparisons between flow stresses in cutting.

Impact Torsion Tests

An impact torsion test machine (Fig. 11) was specially designed with the function of fracturing a specimen by torsion in a very short time at a constant strain rate. The impact energy was stored in a fabricated flywheel (A) which was driven by a d-c motor to any speed up to 950 rev/min. It was so designed that the energy dissipated for breaking the specimen amounted to only 1.6 per cent of the total energy stored in the flywheel running at the highest speed and to 17.5 per cent at the lowest speed used.

The hollow shaft (C) on which the flywheel was keyed was supported by a pair of ball-bearings (B) at the outside diameter and fitted with a small roller-bearing (D), and a cast-iron bushing (G) in the bore, thus enabling its free rotation on top of the solid shaft (E). The solid shaft was held to the specimen (I) and drove it when the clutch was engaged.

To achieve sudden engagement of the driving and the driven, a "solid drive" clutch, similar to the mechanism in a punching press, was developed. Its principal part was a spring-loaded rotating key (L) with an "almond" section which was held by a pin (F) against the spring force into the "home" position as shown when the clutch was not engaged. During the operation the pin was suddenly pulled out by hand and the almond section rotated counter-clockwise to engage the recess at the bore of the hollow shaft which was running at required speed, thus achieving solid engagement of these two shafts.

The angle of twist of the working section of the specimen was measured by the relative movement of two concentric semicylindrical condenser plates (H) fastened to the specimen (I) at different sides of the working section. A Fielden "proximity meter" was used to record the capacitance change. Torque was measured by a torque meter (J) with four strain gages cemented at 45 deg to the axis and forming a Wheatstone bridge. The output of the strain-gage bridge, together with the output of the proximity meter, were recorded by a double-beam d-c amplified cathode ray oscilloscope with photo-recording attachment. For each test the torque-time curve (the single line trace) and the twist angle-time curve (the envelope formed by carrier waves) were superimposed on the oscillogram (Fig. 12).

The specimens (Fig. 13) were of a similar type as for the static torsion tests. A U-slot was cut to form the working section in the middle of the specimen. To avoid possible buckling during the twist and to obtain high strain rates, the working section was made as short as possible. A steel mandrel was also used to increase the stability of the working section.

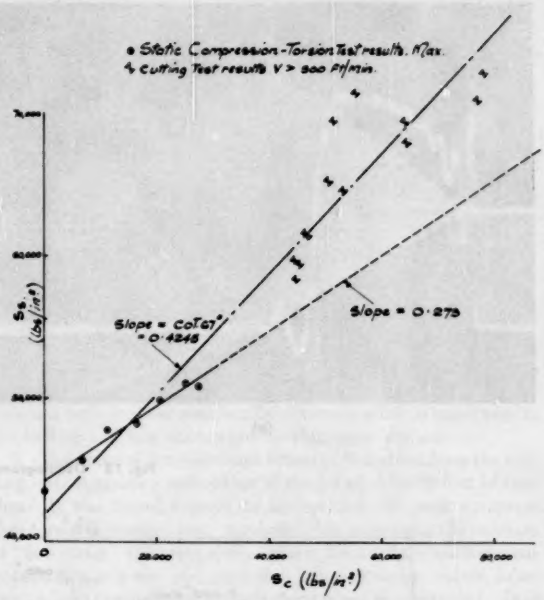

• Static Compression-Torsion Test results. Max.
4 Cutting Test results V > 500 ft/min.
Slope = cot 67° = 0·4245
Slope = 0·273
S_s (lbs/in²)
S_c (lbs/in²)

Fig. 10 Shear stress versus compressive stress

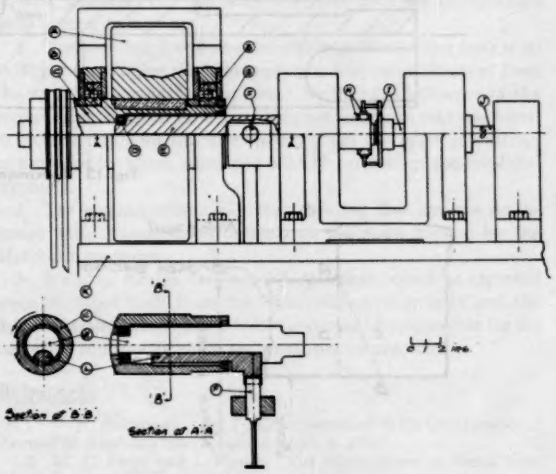
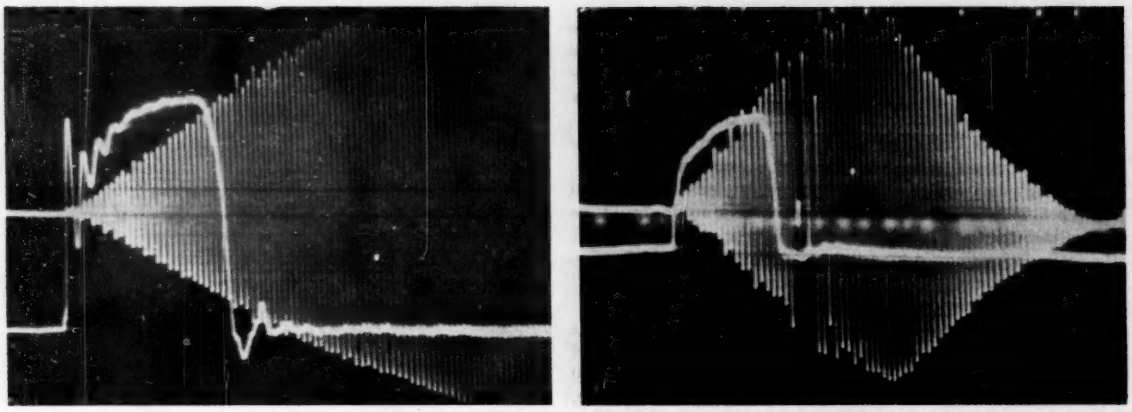


Fig. 11 Impact torsion test machine

Some of the oscillograms obtained showed a fluctuation of stress immediately after the impact as shown in Fig. 12(*a*). In measuring the stress, an approximate mean value was plotted through the initial cycles of the stress fluctuation. To check the validity of this mean plot, a special test was carried out where the strain gages were stuck onto the fixed side of the specimen which was then "built in" as solidly as possible to the machine frame. Hence this section of the specimen was actually used as a torque meter with a very high stiffness. Consequently, a very smooth torque-time curve with insignificant fluctuation was recorded as in Fig. 12(*b*). The stress-strain curve derived from this test was compared with the mean plot of the result of another test under identical conditions but using the separate torque meter, as previously described, and good agreement was found. Thus it was



(a) (b)

Fig. 12 Oscillograms from impact torsion tests

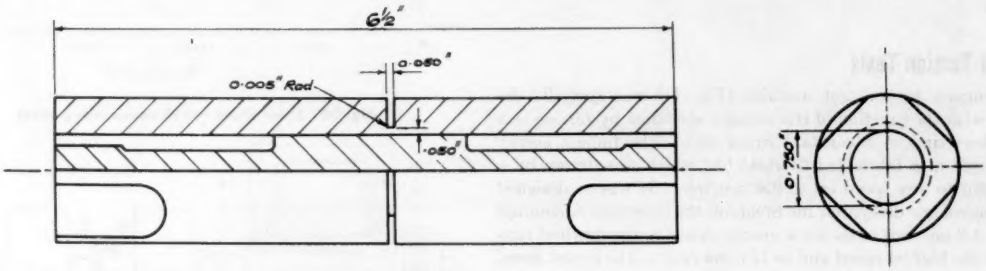


Fig. 13 Specimen for impact torsion tests

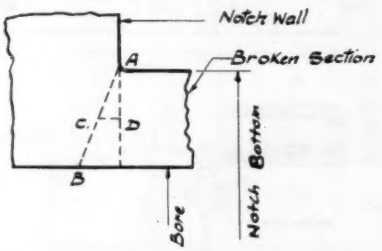


Fig. 14 Spreading of plastic deformation in impact torsion specimens

concluded that the periodic motion of the less stiff torque meter did not affect the type of torque-twist curve obtained. All the remaining tests were carried out using the separate torque meter.

It was found from the specimens tested that some plastic deformation had spread into the main body beyond the fillets. The photomicrographs of the section along the axis of the specimen showed that the spreading was more or less limited to a boundary line AB as shown in Fig. 14. It was also found from these photographs that, in general, the spreading extended deeper in the lower strain-rate tests and nearly disappeared at the highest strain rate, although no definite relationship could be established between the depth of the spread and the strain rate. This is due to the fact that, at higher strain rates, the material exhibits flatter stress-strain curves approaching the ideal plastic elastic form (Fig. 15). In other words, the difference between the initial yield and the ultimate stresses decreases with strain rate and therefore less plastic flow occurs beyond the notch. The strain values were ac-

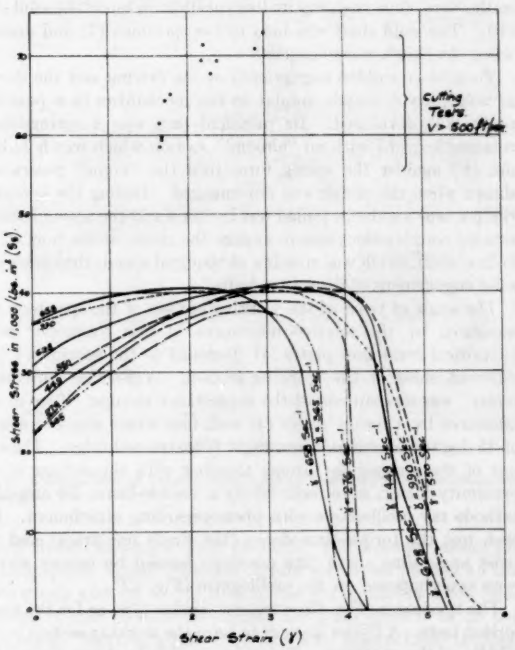


Fig. 15 Comparison of results from cutting and impact torsion tests

cordingly corrected by including the mean depth of spread CD in Fig. 14 to the width of the notch section.

For a long specimen subjected to dynamic plastic deformation, the time taken by the stress wave to propagate along the specimen may be comparable to the total loading time. In such cases the stress recorded at one end of the specimen is not the actual value existing in the other part of the specimen. However, in the present tests, the working section which was the only part undergoing plastic deformation was only 0.05 in. long. The distance from the working section to the torque meter was 7 in. An elastic torsional stress wave would take about 0.06 msec to travel this distance. At the highest speed the whole loading time was 6 msec. With the ratio of the latter to the former to be about 100, the stress and strain recorded should give a reasonable true picture of what is happening in the working section.

Comparison Between the Cutting Tests and the Impact Torsion Tests

The cutting-test results for cutting speeds above 500 ft/min, together with the stress-strain curves obtained from the impact torsion tests at strain rates ranging from 276 to 990 per second, are plotted in Fig. 15. It shows that the dynamic shearing flow stress in cutting is about 40 to 80 per cent higher than the highest stress obtained from the impact-torsion tests. It also reveals that this material does not exhibit any significant change in the ultimate strength, while the yield stress increases considerably with the increasing strain rate. When these curves are compared with the static stress-strain curve without compressive load in Fig. 8, it can be seen that the ultimate strength is about the same but the yield stress is lower in the static tests. Insignificant change of the ultimate strength for alpha-brass at wide range of strain rate was also found by another investigator [11].

The strain rate in metal cutting has been estimated, without experimental evidence, by a few investigators to be within 10^3 to 10^6 per second. Kececioglu [7] was the first one to estimate experimentally the strain rate in the cutting processes. By measuring the width of the "shearing zone" from chips obtained from abrupt stopping tests, he found the value of the strain rate in cutting steels to be from 2500 to 212,000 per second. To the best of the authors' knowledge, no experimental determination of the strain rate in cutting brass has ever been attempted. The highest strain rate achieved in the present impact torsion tests, i.e., 990 per second, is lower than the lowest value determined by Kececioglu for steel. However, judging from the trend of change of the stress-strain curves shown in Fig. 15, further increase in the strain rate in the impact torsion test would produce flatter curves but any significant increase of the ultimate strength would be unlikely.

Shaw, et al. [2] proposed the following relationship to account for the difference between the shear flow stresses in cutting and in material tests:

$$S_s = S_s' + D\gamma + \frac{B}{t_s}$$

where

S_s = shear flow stress in cutting
S_s' = yield stress in torsion test
D = work-hardening modulus in tensile test
γ = shear strain in cutting
B = a constant which was found to be 23 lb/in. for some steel and copper
t_s = length of shear plane extending from the tool tip to the free-chip surface = $t/\sin\phi$

D was taken from tensile test presumably because no torsion data were available. In the present investigation the shear flow stress in torsion ($S_s' + D\gamma$) could be read off directly from the stress-strain curve obtained for any particular value of γ. A sample calculation in the present results showed that $S_s - (S_s' + D\gamma)$ was equal to 20,500 lb/in.^{2} when t_s was 0.0074 in. Using B = 23 lb/in., as found by Shaw, the term B/t_s would account for only 15 per cent of the total discrepancy.

An attempt was also made to find a constant B possibly existing in the present results by plotting $S_s - (S_s' + D\gamma)$ against t_s. The plotting showed the points scattering at random and no consistent value of B could be determined.

Conclusions

1 The Merchant constant C for α-brass increases systematically with increase of the cutting speed until a constant value (67 deg ± 5 per cent) is attained at 500 ft/min and above. Photomicrographs of chips from tests abruptly stopped at about 500 ft/min show good single shear plane characteristics. At lower speeds a definite shear zone can be observed which is narrowest at the tool tip and fans out toward the chip upper surface.

2 The value of the Merchant constant C derived from the cutting tests suggests a high value of the internal coefficient of friction. It was found difficult to account for this from compression-torsion investigation. A relationship expressing the increase in "flow stress" at a *given strain value* in terms of the normal component of stress was not possible. Ignoring strain values, however, a linear increase in *ultimate* shear stress was recorded. It is not at all clear whether this latter procedure is justified; even if it were, however, the effects do not agree with the implications from cutting tests.

3 The shearing flow stress derived from the cutting tests is 40 to 80 per cent higher than the corresponding value observed from the static and impact torsion tests. Neither the influence of the compressive stress on the shear plane nor the strain rate was found to be responsible for this discrepancy. The size effect in cutting, as proposed by Shaw, explained only 15 per cent of the total discrepancy.

4 The cutting tests imply that shearing flow stresses do increase with normal stress but not on the scale implied by the Merchant constant.

5 Shearing strains very much larger than would be expected were produced both from the static compression tests and the dynamic tests, hence these two factors could be responsible for the fact that continuous chips are produced in practice.

References

1 P. W. Bridgman, "On Torsion Combined With Compression," *Journal of Applied Physics*, vol. 14, 1943, p. 273.

2 M. C. Shaw and I. Finnie, "The Shear Stress in Metal Cutting," TRANS. ASME, vol. 77, 1955, p. 115.

3 M. E. Merchant, "The Mechanics of Metal-Cutting Processes," *Journal of Applied Physics*, vol. 16, 1945, pp. 267 and 318.

4 B. T. Chao and G. H. Bisacre, "The Effect of Speed and Feed on the Mechanics of Metal Cutting," *Proceedings, I. Mech. E.*, vol. 165, 1951, p. 1.

5 B. T. Chao, K. J. Trigger, and J. B. Zylstra, "Thermophysical Aspects of Metal Cutting," TRANS. ASME, vol. 74, 1952, p. 1039.

6 P. Bastien and M. Weisz, "Experimental Study of the Mechanism of Plastic Deformation in Metal Cutting Compared With the Methods of Classical Stressing," *Microtechnic*, vol. 11, 1957, pp. 122 and 151.

7 D. Kececioglu, "Shear-Strain Rate in Metal Cutting and Its Effects on Shear-Flow Stress," TRANS. ASME, vol. 80, 1958, p. 158.

8 H. Ernst, "Physics of Metal Cutting," *Machining of Metals*, 1938, p. 24.

9 W. B. Heginbotham, "An Experimental Investigation Into the Characteristics and Behavior of the Built-up Nose When Machining Mild Steel With High-Speed Steel Tools," *Microtechnic*, vol. 11, no. 3, 1957.

10 S. S. Gill, "Three 'Neutral' Loading Tests," *Journal of Applied Mechanics*, vol. 23, TRANS. ASME, vol. 78, 1956, p. 497.

11 H. G. Baron, "Stress-Strain Curves of Some Metals and Alloys at Low Temperatures and High Rates of Strain," *Journal, Iron and Steel Institute*, vol. 182, 1956, p. 355.

S. KOBAYASHI[2]
Assistant Professor of Mechanical Engineering, Doshisha University, Kyoto, Japan

E. G. THOMSEN
Professor of Metal Processing, University of California, Berkeley, Calif. Mem. ASME.

The Role of Friction in Metal Cutting[1]

Metal-cutting studies were made with free-cutting steel SAE 1112 and alloy steel SAE 4135 in the as-received condition with artificially controlled tool-chip contact areas and flank contact areas (artificial wear lands). The experimental results for steel SAE 1112 at a speed range of 0.083 to 1010 fpm, reveal that friction under metal-cutting conditions on the rake face can be explained satisfactorily by a junction model with possible superimposed general plastic flow above the junctions in accordance with the general rules of plastic deformation (von Mises instantaneous yield criterion). The experimental results also reveal that the friction mechanism at the controlled flank-wear contact area is essentially the same as that occurring at the tool face. The difference in behavior of the two steels was attributed to the role that manganese sulfide appears to play in free-cutting steels.

THE subject of friction is important in all engineering applications wherever solid surfaces are in sliding contact with each other. This is particularly true in metal-working processes where the sliding pair of surfaces are metals and where plastic deformation of the softer of the two metals usually takes place under conditions of high normal pressure. It is, therefore, of importance to understand the mechanics of sliding friction in order to reduce the resulting drag to as low a value as possible if it happens to be present solely as a useless energy-consuming mechanism. This is the case with metal cutting and it should not be too surprising to find that the friction process in metal cutting has received appreciable study.

In orthogonal metal-cutting processes frictional drag apparently is encountered on the rake face of the tool between chip and tool and, as the tool wears, additional frictional drag also occurs between the flank of the tool and the workpiece. The size of the area of contact of these rubbing surfaces is determined by the cutting processes and, therefore, contact takes place over a length known as the natural length on the face of the tool for the particular cutting condition under investigation as well as the wear land on the flank of the tool. The two frictional processes resulting from contact at these surfaces, however, are difficult to study under ordinary cutting conditions, since both normal pressure and shear stress of the work material are not readily determinable. It is for this reason that recently a number of investigators have limited the extent of these contact surfaces artificially in order to study the frictional processes under metal-cutting conditions. Thus Takeyama and Usui [1][3] studied free-cutting brass and found that both normal and shear forces decreased linearly with decreasing contact length on the rake face of the tool. These observations appear to be at variance with the results reported by Chao and Trigger [2] for steel AISI 4142 and electrolytic copper, and by Kobayashi, et al. [3] for free-cutting steel SAE 1112. Takeyama and Usui reported, as did Chao and Trigger, that the frictional processes on the rake face of the tool involve bulk metal shearing of the chip material as it slides over the known contact length between chip and tool and that the ratio of real area of contact to apparent area of contact may approach unity under certain cutting conditions. By real area of contact in this case is meant the actual contact area between asperities of the tool and the chip, while apparent area is equal to the nominal contact area determinable from the artificially controlled contact length. Other experimental results using controlled wear lands were reported by Olberts [4], Okoshi and Sata [5], and Kobayashi and Thomsen [6]. While Olberts' principal aim was to study the increase in average tool temperature as measured by a work-tool thermocouple technique, the other investigators [5, 6] concentrated their efforts on studying the frictional processes involved in the rubbing of these surfaces. In their investigations Okoshi and Sata found shearing stresses on the wear land approximately equal to those on the rake face when cutting an annealed SAE 4140 steel at approximately 490 fpm. In contrast, Kobayashi and Thomsen found that the shear stresses for a free-cutting steel and an SAE 4135 steel in the as-received conditions were smaller than those on the rake face. This latter observation appears to be in accord with the effects of work-hardening on the shear-flow processes and will be discussed in a subsequent section of this paper.

Experimental Procedure

The experimental procedure of measuring forces and controlling cutting conditions has been reported in detail elsewhere [6, 7]. In order to provide the tools with artificially controlled contact areas, however, for the friction studies on the rake face and flank of the tool, the following procedures were used:

1 The rake face was undercut parallel to, but at an arbitrarily selected distance from, the cutting edge. The tool was then ground at its flank and the resulting distance between the cutting edge and the relief cut on the face was accurately determined with a toolmaker's microscope.

2 The flanks of the cutting tools were first provided with a secondary clearance angle of 5 deg. The controlled contact length was then ground on the flanks to various desired lengths at clearance angles of 0, $-1/2$, and -1 deg. The accuracy obtained in these angles when the tools were finally adjusted in the dynamometer was of the order of better than ± 5 min. It was estimated that the tendency of the moments, thus produced by the force system at the flank, to rotate the tool and thus alter this angle was nil.

All tools for the low-speed cutting tests at cutting speeds of 100 fpm or less were high-speed steels of the 18-4-1 (Rex AA) type,

---

[1] This paper is the result of an investigation sponsored by the Machinability Program located at Rock Island Arsenal under Contract DA-11-070-508-ORD-605 with the University of California.

[2] At present, graduate student in Mechanical Engineering at the University of California, Berkeley, California.

[3] Numbers in brackets designate references at end of paper.

Contributed by the Metal Processing Activity of the Production Engineering Division and presented at the Annual Meeting, Atlantic City, N. J., November 29–December 4, 1959, of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS.

NOTE: Statements and opinions advanced in papers are to be understood as individual expressions of their authors and not those of the Society. Manuscript received at ASME Headquarters, June 16, 1959. Paper No. 59—A-132.

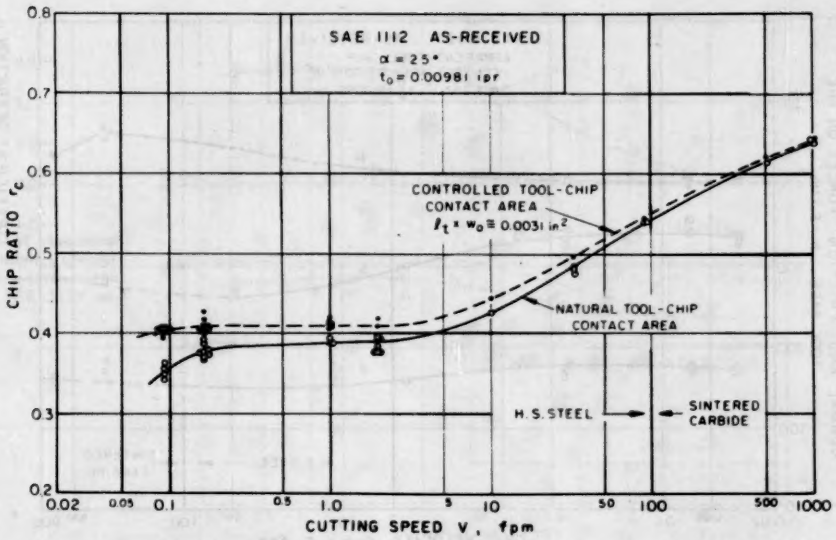


Fig. 1 Comparison of chip ratios for natural and controlled tool-chip contact areas for steel SAE 1112, as-received, at various cutting speeds

while those for the higher speeds were sintered-carbide tools (K3H).

Controlled Tool-Chip Contact-Area Tests

In order to study the friction process on the rake face of the tool, free-cutting steel SAE 1112 in the as-received condition of Rockwell hardness C8 was cut dry by end-cutting of tubular specimens at speeds ranging from 0.083 to 1010 fpm. The method of testing under these conditions is described in another paper [7] and required a special technique for the low-speed tests. The cutting tools were provided with true rake angles of 25 deg and flank clearance angles of 3 deg. The feed was held constant at 0.0098 ipr. The results of chip ratio as a function of cutting speed for natural contact and restricted contact lengths are shown in Fig. 1. The restricted tool-chip contact area was 0.0031 sq in. (contact length l_t = 0.0155 in. and width of cut w_0 = 0.200 in.), which was chosen to yield a contact length slightly smaller than the natural contact length for the whole speed range in order that the so-called apparent contact area be known. It is seen in the figure that this condition is fulfilled since the tests with natural tool-chip contact lengths gave chip ratios which were slightly lower for the whole speed range than those with restricted contact lengths. It is at once evident from the figure that speed has a profound effect on the magnitude of the chip ratio as the speed increases beyond approximately 2 fpm. Since it is known now that the shear process on the shear plane is unaffected by speed [3, 6, 7], it must be concluded that the rising chip ratio at the higher speeds is principally due to a reduction in the frictional drag at the tool-chip interface.

It is now of interest to compare the forces acting on the controlled tool-chip contact area of 0.0031 sq in. The results are given in Fig. 2 and it is seen that the normal force N drops more rapidly with speed than does the friction force F. As a consequence of this, the coefficient of friction μ increases with speed except at the highest speed employed, in spite of the fact that the cutting condition is improving. This observation emphasizes again the fact that a coefficient of friction under the severe conditions of pressure, temperature, and speed as found in metal cutting becomes meaningless and may lead to misinterpretation.

Before considering the nominal stresses acting in the chip at the tool-chip interface which are now calculable for the known apparent area A_a, it will be of interest to review present thought on sliding friction between dry metallic surfaces. In a recent paper Finnie and Shaw [8] reviewed the literature on dry-friction studies and gave credit to several investigators, namely, Holm [9], Ernst and Merchant [10], and Bowden and Tabor [11], for suggesting the composite mechanism which they believe it to be. This composite mechanism of friction consists of the following three factors:

1 Mechanical interlocking of surface asperities.

2 Plowing of asperities of the harder surface through the softer of the two sliding metals.

3 Formation of metallic junctions by welding of asperities.

It does not seem possible to say, a priori, which one of these factors contributes the most to the resistance to motion, when a ductile metal slides over a relatively hard surface. This lack of detailed knowledge has led Green [12] to investigate the foregoing effects with plasticine models. Green found that the theoretical junction strength depends on the type of junction, i.e., strong or weak, and that for strong junctions the coefficient of friction $\mu = \delta F/\delta N$ must be between $0.4 < \mu < +1.0$, where δF and δN are the shear and normal forces, respectively, for any particular junction. Green did not consider work-hardening, stress concentration, size effect, or others in his analysis, and concluded that $\delta F/\delta A = \tau$, the maximum shear stress for any junction, can be somewhat greater than the instantaneous bulk shear strength of the metal. It also appears from Green's work that there is no substantial difference between the three mechanisms just given, and that it may perhaps be possible to treat them as one and the same process in estimating frictional drag.

Finnie and Shaw [8] take a different view in that they maintain that the stress to break a junction is much higher than the bulk yield shear strength of the metal. They consider welded junctions only and base the high strength of a junction on the fact that its area may be extremely small and hence the shear stress necessary to rupture this junction may not be that which can be determined by ordinary rules of plastic deformation. These authors assume that the size of the junction may in fact be so small that insufficient dislocations are present in the soft metal of the junction in order to initiate plastic deformation and consequently the

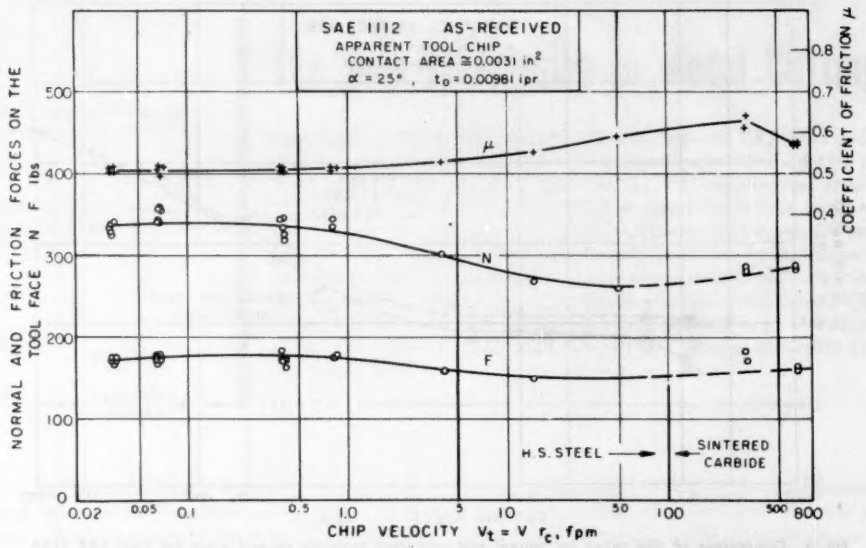

SAE 1112 AS-RECEIVED
APPARENT TOOL CHIP CONTACT AREA ≅0.0031 in²
α = 25°, t₀ = 0.0098 ipr
NORMAL AND FRICTION FORCES ON THE TOOL FACE N F lbs
COEFFICIENT OF FRICTION μ
μ
N
F
H.S. STEEL
SINTERED CARBIDE
CHIP VELOCITY Vt = V rc, fpm

Fig. 2 Normal force N, friction force F, and coefficient of friction μ on tool face as functions of chip velocity for steel SAE 1112, as-received, with controlled tool-chip contact areas

stress necessary to rupture a junction may in the limit approach the theoretical shear strength of the crystal. This shear strength is given as $\tau = G/2\pi$, where G is the modulus of rigidity. The magnitude of this limiting shear strength depends on the elastic properties of the metal and is approximately 2×10^6 psi for steels which is of the order of 20 to 40 times as large as the bulk shear strength observed in cutting tests with ordinary steels [6, 7]. This effect has been called a "size effect" and has been used by Shaw, et al. [13] to explain in part the role of lead in metal cutting and by Shaw and Finnie [14] to explain the high chip loads in metal cutting at small feeds.

According to Bowden and Tabor [11], the equation, applicable to dry friction for each welded asperity can be written as follows:

$$\mu = \frac{\delta F}{\delta N} = \frac{\delta A_r \tau}{\delta A_r p} = \frac{\tau}{p} \tag{1}$$

where τ and p are the shear yield strength and yield pressure for the softer metal, respectively, and δA_r is the area of the junction. Finnie and Shaw [8] rewrote this equation in the form

$$\mu = \frac{KA_r\tau^*}{A_r p^*} = K\frac{\tau^*}{p^*}, \tag{2}$$

in which KA_r is the fraction of the asperities in contact with one another which have welded to each other. The magnitude of K becomes unity, when all contacting asperities have formed welded junctions. The values τ^* and p^* have a similar meaning as τ and p of equation (1), except that the yield strengths of the asperities are considered to be much higher than the so-called bulk yield strength of the softer metal.

Let us now assume that K of equation (2) is unity and compare equations (1) and (2) with the experimental results of the present investigation. We may do this by assuming that the junctions are numerous enough to lock the whole surface of the sliding chip to the rake face of the tool. Then relative motion may not necessarily take place at all junctions, but instead bulk flow may take place in a layer in the chip parallel to the tool face, but some distance above the junctions. This hypothesis is in part supported by the fact that highly deformed grains can readily be observed in photomicrographs of sectioned chips on the tool side of the

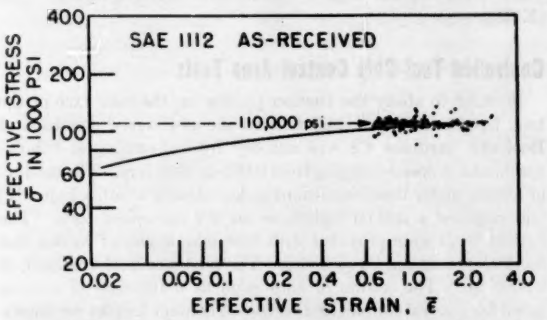


Fig. 3 Correlation of metal-cutting data with compression-test data for steel SAE 1112, as-received, based on distortion-energy theory

chip. Thus we may designate the bulk flow area as A_b, which may be much larger than the real area A_r of Finnie and Shaw [8]. In fact, let us assume for the moment that the bulk flow area A_b is equal to the apparent area A_a. If this is the case we may take the data in Fig. 2 at low cutting speeds and divide the forces by the apparent area ($A_a = 0.0031$) in order to obtain bulk flow stresses. This is possible, however, only then if all measured dynamometer forces are transmitted through the chip. For the case of free-machining steel (SAE 1112 as-received) the so-called tool-edge force [1, 6, 7] is small and may be neglected in comparison with the chip forces, when the feed = 0.0098 ipr, as is the case with the forces given in Fig. 2.

If it is assumed now that the metal in the chip is at the yield condition under a simple state of compression, then the instantaneous yield flow stress must be equal to the effective stress $\bar{\sigma}$. The magnitude of $\bar{\sigma}$ is given in Fig. 3 and is approximately 110,000 psi. The foregoing hypothesis requires the assumption that the metal in the chip is uniformly work-hardened to the state of the metal in the shear plane and is not work-hardened further when sliding over the tool face. This latter condition is not necessarily fulfilled, but the effect of further work-hardening has little influence on the average flow stress in view of the fact that the slope of the stress-strain curve at the metal-cutting strains, as seen in Fig. 3, is relatively small. Thus, at the low cutting

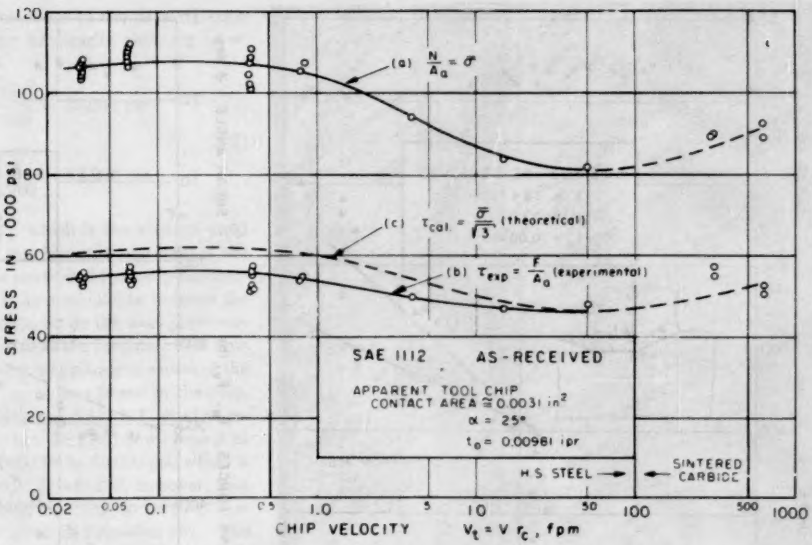

STRESS IN 1000 psi
(a) $\frac{N}{A_a} = \bar{\sigma}$
(c) $\tau_{cal} = \frac{\bar{\sigma}}{\sqrt{3}}$ (theoretical)
(b) $\tau_{exp} = \frac{F}{A_a}$ (experimental)
SAE 1112 AS-RECEIVED
APPARENT TOOL CHIP CONTACT AREA ≅ 0.0031 in²
α = 25°
t_o = 0.00981 ipr
H.S. STEEL
SINTERED CARBIDE
CHIP VELOCITY $V_t = V r_c$, fpm

Fig. 4 Comparison of theoretical and experimental frictional stresses in controlled tool-chip-contact-area studies with steel SAE 1112 as-received

velocities ($V \leq 10$ fpm) where the temperature in the chip remained essentially at room temperature the instantaneous bulk-yield-flow stress (as seen in Fig. 4) is approximately 110,000 psi. Hence,

$$\bar{\sigma} = \frac{N}{A_b} = p_y \tag{3}$$

(see equation 1) is satisfied for the low-cutting-speed tests in Fig. 4. If it is assumed that the plasticity condition in the chip near the tool interface is such that the normal force N produces a state of hydrostatic pressure because of the restraint induced by the asperities, then plastic deformation by superposed shearing will take place only then, whenever

$$\bar{\sigma} = \sqrt{3}\ \tau_{calc} \tag{4}$$

Thus the required calculated shearing stress for the low-speed tests should be

$$\tau_{calc} = \frac{\bar{\sigma}}{\sqrt{3}} = \frac{110{,}000}{\sqrt{3}} = 63{,}500.$$

This value of τ_{calc} checks with the shear-flow stress on the shear plane for this material [6], and also checks closely with the values of τ_{exp} observed on the rake face in Fig. 4. If it is assumed that the normal stress $\bar{\sigma}$ varies due to the combined effect of strain rate and temperature with increasing speed in accordance with the upper curve in Fig. 4, then the shearing stress can be calculated by equation (4) for all speeds. It is seen that the experimental and theoretical values correlate well and give support to the hypothesis that $A_b \cong A_a$ and the metal in the chip near the asperities must flow in bulk. These observations appear to be substantiated by the data of Chao and Trigger [2].

It should be noted, however, that the foregoing results in no way invalidate the contention of Finnie and Shaw [8], that the total area of welded junction A_r is less than the apparent area A_a and that the stresses to break a junction may be much higher than that given by the bulk flow properties. Indeed, that this mechanism correctly describes the process appears to be the only satisfactory explanation why grid marks or tool marks persist on flat surfaces, when the surface is deformed by an indenter, or the reason that it is so easy to dislodge a built-up edge from a tool by a relatively small load.

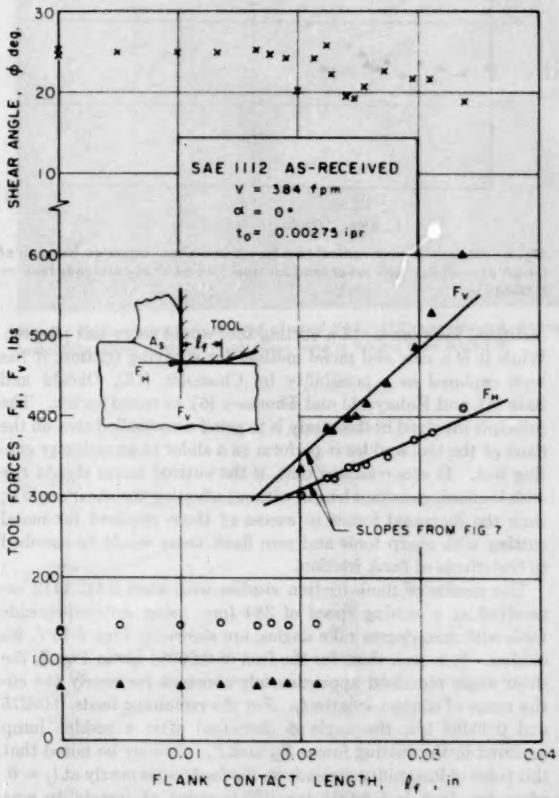


Fig. 5 Horizontal and vertical tool forces and shear angle as function of length of artificial flank wear land for steel SAE 1112, as-received; feed = 0.00275 in.

Controlled-Flank-Area Tests

In studying friction under high pressure, it is important to have a rigid setup, uncontaminated surfaces, and sliders of known areas. It was thought that controlled contact areas at the flank

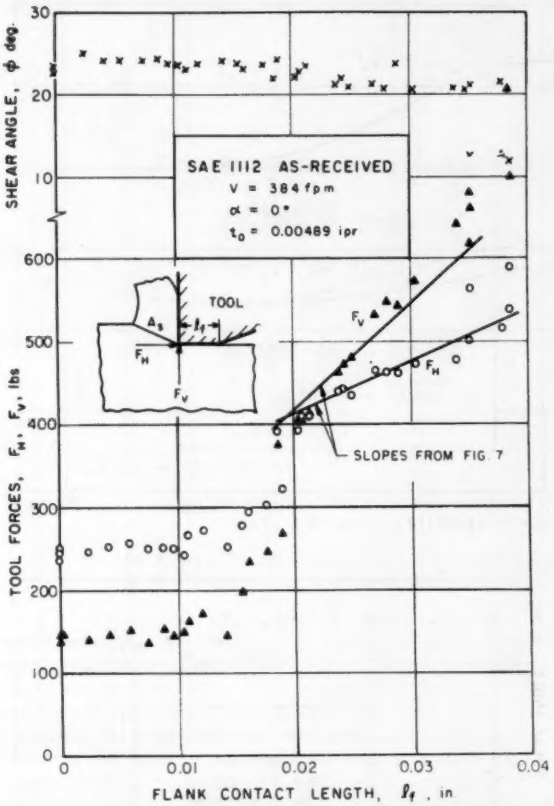


Fig. 6 Horizontal and vertical tool forces and shear angle as function of length of artificial flank wear land for steel SAE 1112 as-received; feed = 0.00489 in.

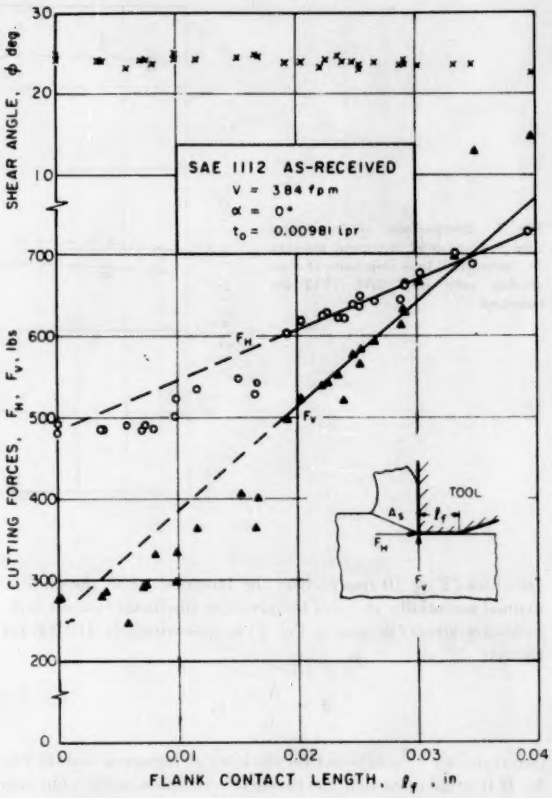


Fig. 7 Horizontal and vertical tool forces and shear angle as function of length of artificial flank wear land for steel SAE 1112, as-received; feed = 0.00981 in.

(artificial wear lands) of a cutting tool would serve this purpose. While it is a new and novel method for studying friction, it has been explored as a possibility by Chisholm [15], Okoshi and Sata [5], and Kobayashi and Thomsen [6], as noted earlier. The principle involved in this study is to grind a controlled area on the flank of the tool and let it perform as a slider in an ordinary cutting test. It was reasoned that, if the cutting forces should rise with increasing contact length without affecting the shear angle ϕ, then the increased forces in excess of those required for metal cutting with sharp tools and zero flank areas would be ascribed to the effects of flank friction.

The results of flank-friction studies with steel SAE 1112 as-received at a cutting speed of 384 fpm, using sintered-carbide tools with zero-degree rake angles, are shown in Figs. 5 to 7, inclusive. It is seen that, for the feed of 0.00981 ipr in Fig. 7, the shear angle remained approximately constant for nearly the entire range of contact lengths l_f. For the remaining feeds, 0.00275 and 0.00489 ipr, the angle ϕ decreased after a sudden jump occurred in the cutting forces F_H and F_V. It may be noted that this point of instability depends on the feed and is nearly at $l_f = 0$, when the feed is 0.00981 ipr. This point of instability was thought to be due to a sudden puncturing of the low-shear-strength films which are present at the interface due to manganese-sulfide inclusions. This film apparently prevents the asperities from contacting at first, but, as rubbing due to increased contact length l_f becomes more severe, the surface temperature of the workpiece increases apparently sufficiently to permit additional welding of asperities.

A preliminary survey with an imbedded thermocouple in the workpiece surface indicated that the local surface temperature for a contact length of $l_f = 0.017$ in. and a feed of 0.00489 ipr, Fig. 5, was 350 F and higher. No temperature readings for large contact areas are presently available, but it must be assumed that the workpiece surface temperature must rise appreciably when the length l_f is such that local instability is initiated, since for thin chips (as for example feed = 0.00275 ipr) the shear-plane angle ϕ suddenly decreases. The fact that the shear angle ϕ decreases for steel SAE 1112 as the workpiece temperature increases will be discussed later since it is opposite in trend to that observed with steel SAE 4135. It may be noteworthy to observe that the forces after the sudden jump rise approximately linearly with increasing flank contact lengths (as indicated by the solid lines), thus suggesting a similar friction mechanism for all feeds.

We may now inquire if the friction criterion discussed earlier is still operative. Inasmuch as the rake angle was zero ($\alpha = 0$), the total forces F_H and F_V can be reduced by the amount required for chip deformation by an intercept correction. The corrected flank forces then are N_f and F_f, provided the shear angle remains constant. Thus, using the data illustrated in Fig. 7, for the feed of 0.00981 ipr, the normal force N_f and shearing force F_f which are operative on the flank surface will be

$$F_f = F_H - 485$$

and (5)

$$N_f = F_V - 275$$

Thus the nominal flank stresses can be calculated by dividing the forces of Equations (5) by the flank area, assuming that the

straight lines drawn are fair approximations to the data. Hence the following results are obtained by arbitrarily choosing l_f = 0.034 in:

$$\tau_f = \frac{F_f}{A_f} = \frac{695 - 485}{0.034 \times 0.200} = 31{,}000 \text{ psi}$$

$$\bar{\sigma}_f = \frac{N_f}{A_f} = \frac{695 - 275}{0.034 \times 0.200} = 62{,}000 \text{ psi} \quad (6)$$

It is seen from these results that $\bar{\sigma}_f$, which is the average yield flow stress at the interface, is now roughly only one half of the magnitude of that encountered in the controlled tool-chip-contact-area experiments, Fig. 4. But this is as it should be because the chip is in a highly work-hardened state, while the workpiece surface is essentially in a virgin state. Hence the instantaneous flow stress required on the flank must have a magnitude in excess of the initial yield stress, but appreciably below that found in the chip. Referring to the compression properties of SAE 1112 steel as-received, given in Fig. 3, it is evident that the yield flow stress $\bar{\sigma}$ at low strains should be of the order of 60,000 to 65,000 psi, which is in agreement with $\bar{\sigma}_f$ in Equation (6). It is found, however, that τ as calculated by Equation (4) would turn out to be $\bar{\sigma}/\sqrt{3}$ = 36,000 psi instead of 31,000 psi as given in Equation (6). The magnitude of the stresses appears to show that now the bulk area A_b is slightly smaller than the apparent area A_a. Hence the experimental data seem to indicate that the friction processes for steel SAE 1112 on the tool face and on the flank area are essentially of identical nature.

It is of interest to note that a second instability for all feeds occurs at all speeds at a flank wear land $l_f \cong 0.030$ in. This instability appears to affect mainly the normal force and is associated with sudden alternating increases and decreases of chip thickness as shown in Fig. 8. This phenomenon has not been explained completely, but undoubtedly is due to a change in the friction mechanism at the flank, which affects the temperature and force system in the chip. It may be argued, for example, that rupturing of junctions is the primary process of friction before the point of instability is reached, which would result in only a nominal temperature rise in the workpiece surface. If suddenly a large number of junctions weld to the tool, due to a slight increase in temperature, it may result in a sudden shift of the shear mechanism from the junctions to a layer above the junctions, which in turn, could set off a series of events, such as inducing high surface temperature in the work piece, followed by a change in friction mechanism there, both of which would affect the shearing process and friction in the chip.

In order to check whether or not the flat portion of the force versus flank contact-length relationship, shown in Fig. 5, may be due to some misalignment or tool distortion which could possibly cause the clearance angle to change a few minutes during cutting, the sintered-carbide tools were provided with negative wear-land-clearance angles χ_w of appreciable size, namely, $-1/2$ and -1 deg. It was thought that this would insure forced contact at the flank. The results are given in Fig. 9 and it is seen that even the enforced ploughing of the metals at $-1/2$-deg negative clearance angle did not eliminate entirely the lack of abhorrence of the metal to frictional drag at small contact lengths. It is seen further that, as the angle increases to -1 deg, the flat portion disappears, but that the slope of the force versus contact-length curves remains nearly the same, as indicated by the solid lines drawn at constant slope. This appears to indicate that, for steel SAE 1112 under the conditions tested, ploughing, welding, and rupturing of asperities must be closely related to each other in friction processes of sliders at high normal pressures.

Experiments similar to those just described were also carried out with alloy steel SAE 4135 in the as-received condition. The results are given in Figs. 10 to 13, inclusive. This material does not

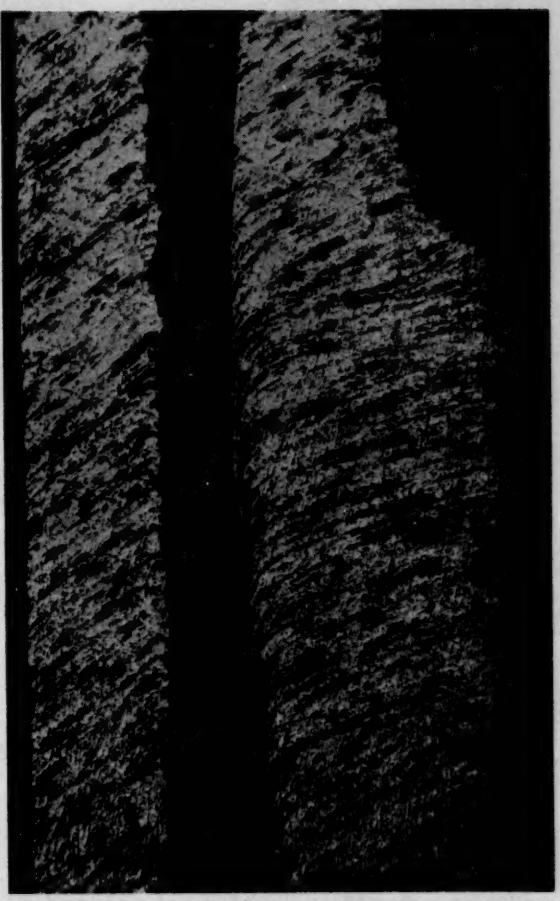


Fig. 8 Photomicrograph of chip of steel SAE 1112, as-received, during instability when artificial flank wear land is longer than 0.030 in.

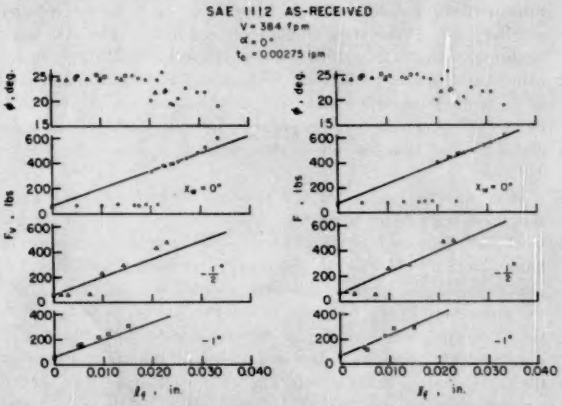


Fig. 9 Horizontal and vertical tool forces and shear angle as function of length of artificial flank wear land for steel SAE 1112, as-received, for flank-wear-land clearance angles of 0, −1/2, and −1 deg

differ too greatly from that of steel SAE 1112 in so far as the plastic-flow properties are concerned, but it lacks the excess of manganese sulfide which distinguishes the free-cutting steels from others in their performance during metal cutting. However, it is seen from the curves in Figs. 10 to 12, that this alloy exhibits a

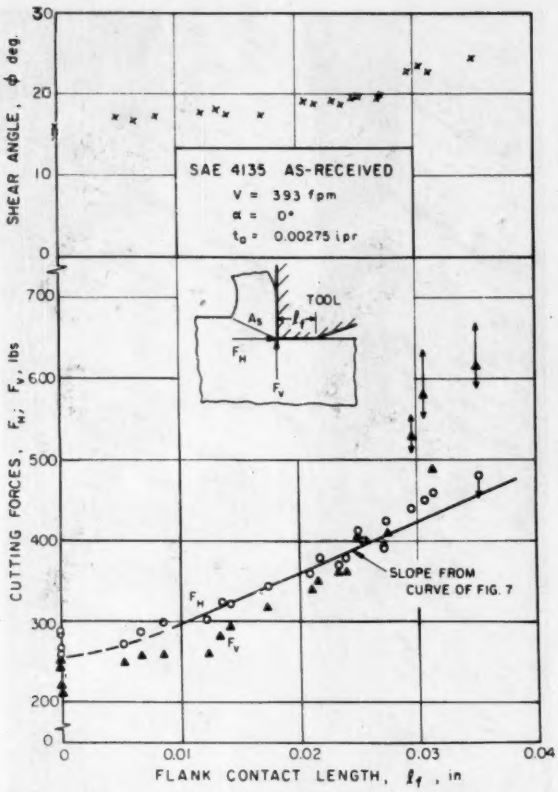


Fig. 10 Horizontal and vertical tool forces and shear angle as function of length of artificial flank wear land for alloy steel SAE 4135, as-received; feed = 0.00275 in.

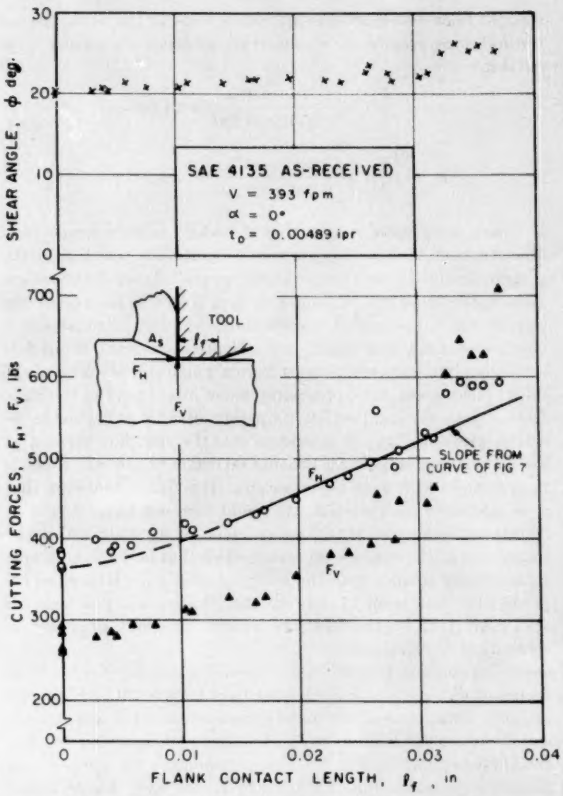


Fig. 11 Horizontal and vertical tool forces and shear angle as function of length of artificial flank wear land for alloy steel SAE 4135, as-received; feed = 0.00489 in.

number of important differences in behavior when compared with steel SAE 1112. These are (*a*) the intercept forces are higher for all speeds, (*b*) the discontinuities at low feeds seem to have disappeared, (*c*) the initial slopes of the F_V versus l_f curves are much smaller, and (*d*) the shear angle increases rather than decreases, with increasing l_f as was the case with steel SAE 1112. On the other hand, the instability in the normal force F_V at magnitudes of $l_f \cong 0.030$ in. still exists and the slopes of the F_H versus l_f curves at larger values of l_f appear to be approximately the same although this trend is not as definite as it was with steel SAE 1112.

It is unfortunate that at this time not all of the foregoing observations can be explained satisfactorily and, therefore, more work must be done before a clear-cut concept of the mechanism of flank friction will emerge. Nevertheless, some tentative explanations can be given at this time in order to attempt to clarify the differences in the behavior of the two steels. The fact that the intercept forces are higher with steel SAE 4135 than with steel SAE 1112 appears to be associated with the singular force at the cutting edge. It is seen in Fig. 14, for example, that the intercept forces for steel SAE 1112 at zero feed and 0-deg rake angle, which are associated with a tool-edge force, are nearly zero at a cutting speed of 384 fpm. In contrast to this, at a cutting speed of 1010 fpm, shown in Fig. 14, the forces are much larger and, obviously, cannot be ignored.

It is thought that under certain conditions the flank might participate in giving added resistance to the tool-edge force in opposing tool motion, and thereby cause the intercept forces to be large. The magnitude of such forces would depend on the cutting condition and no precise estimate is available. One can get an order of magnitude, however, by noting the change in tool forces, shown in Fig. 15, when changing the flank-clearance angle for alloy steel SAE 4135 as-received from 3 to 1 deg. It appears that this increase must be due entirely to flank friction, possibly caused by rapid flank wear, welding of a built-up flank caused by wear particles, and rolling and ploughing of wear particles. In view of this, it appears reasonable to assume that the relatively small rise in F_H with increasing l_f for steel SAE 4135 as-received, at short artificial wear lengths, may be due to the fact that the flank near the tool edge already is in contact with the workpiece. Hence any added flank contact surface does not materially alter the tool forces until the contact length becomes appreciably longer. The fact that the slopes of the F_H versus l_f curves at longer l_f for the two steels tend to be the same gives credence to the hypothesis.

The data discussed in the foregoing may also be used to interpret the role that manganese sulfide may play in the cutting process. It may be remembered that at all speeds with artificially controlled tool-face contact areas, the shear-flow stresses for free-cutting steel SAE 1112, shown in Fig. 4, were nearly equal to the theoretical values, assuming that the friction Equations (3) and (4) are applicable. The implication then is that the apparent area A_a is nearly equal to the bulk area A_b, and the low-shear-strength material in the inclusions of the free-cutting steel is ineffectual in overcoming friction. This is to say that the rubbing pair of tool and chip are in a range which Shaw, et al. [13] would describe as producing catastrophic wear.

Inasmuch as the normal pressure in the chip must gradually

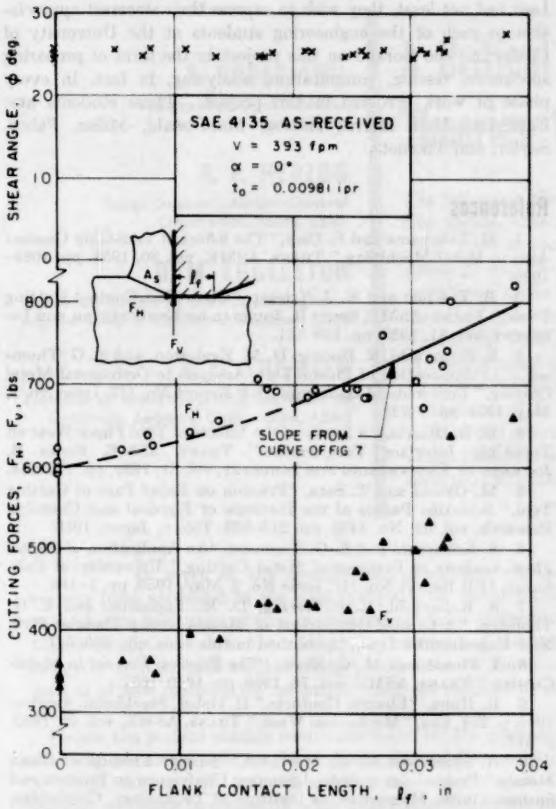

SAE 4135 AS-RECEIVED
V = 393 fpm
α = 0°
t0 = 0.00981 ipr
SHEAR ANGLE, φ deg.
CUTTING FORCES, FH, FV, lbs
SLOPE FROM CURVE OF FIG 7
FLANK CONTACT LENGTH, lf, in

Fig. 12 Horizontal and vertical tool forces and shear angle as function of length of artificial flank wear land for alloy steel SAE 4135, as-received; feed = 0.00981 in.

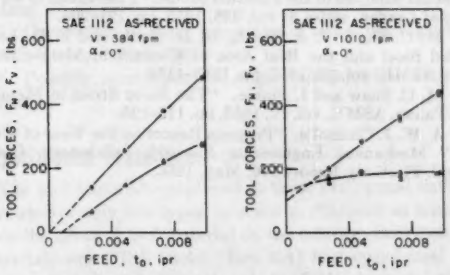


Fig. 14 Cutting forces as function of feed for steel SAE 1112, as-received, at cutting speeds of 384 and 1010 fpm

drop to zero downstream from the cutting edge, it must be assumed that a point will be reached where the normal pressure will be low enough that the manganese-sulfide film can no longer be ruptured. This would have the effect of decreasing the frictional drag to low values even though the normal pressure may still be of appreciable magnitude. Comparing this behavior with that of alloy steel SAE 4135 it could be reasoned that the chip-tool contact length for this alloy would be longer than that for free-cutting steel, since a natural antiwelding, low-shear-strength film is not present. Measurements on contact length have been made and seem to confirm the foregoing contention. Thus the effect would be that for alloy steel SAE 4135 the total contact length is longer, requiring that a greater normal force and a greater friction force be overcome.

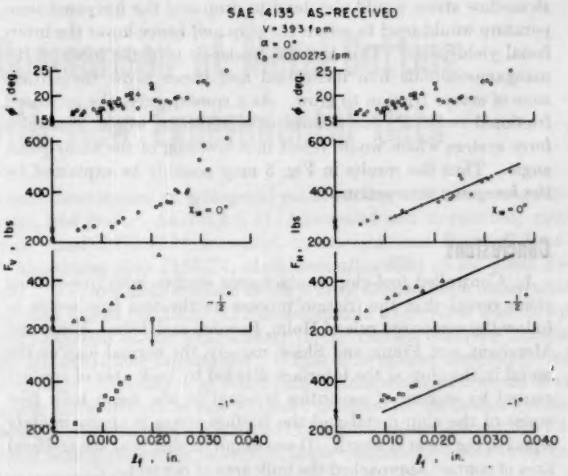


Fig. 13 Horizontal and vertical tool forces and shear angle as function of lengths of artificial flank wear land for alloy steel SAE 4135, as-received, for flank wear-land clearance angles of 0, −1/2, and −1 deg

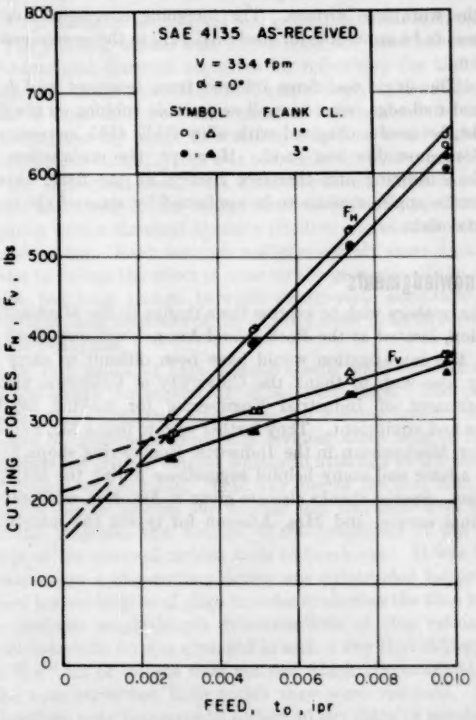


Fig. 15 Effect of flank-clearance angles on the horizontal and vertical tool forces for alloy steel SAE 4135, as-received

In the light of the foregoing argument, we might examine the effect of a preheated work surface as, for example, that preheating which is due to flank friction. For alloy steel SAE 4135 the effect would be that the shear-flow stress in the shear plane would be lowered tending to increase the shear-plane angle, while additional heating of the chip would possibly cause some decrease in friction on the tool face because of the fact that the metal in the tool-chip interface has a lower yield strength. It is possible that the net effect may be to raise the angle ϕ as observed in Fig. 10. When steel SAE 1112 enters the shear zone in a preheated condition, the

shear-flow stress would also tend to drop and the increased temperature would tend to soften the chip and hence lower the interfacial yield point. This has the tendency to make more of the manganese-sulfide film ineffectual and hence allow the critical zone of severe friction to grow. As a consequence, the increased frictional resistance for reasons of equilibrium would demand a force system which would result in a lowering of the shear-plane angle. Thus the results in Fig. 5 may possibly be explained by the foregoing interactions.

Conclusions

1 Controlled tool-chip contact-area studies with free-cutting steels reveal that the friction process on the tool face seems to follow the composite rule of Holm, Bowden and Tabor, Ernst and Merchant, and Finnie and Shaw, namely, the normal load on the metal in the chip at the interface divided by bulk area of contact caused by seizure of asperities is equal to the mean bulk flow stress of the chip metal and the friction stress is approximately equal to the shear strength. It was found further that the artificial area of contact approached the bulk area of contact.

2 Controlled flank contact-area tests (artificial wear lands), like tool-chip contact areas, appear to be suitable for friction studies when the shear angle is not affected by temperature rise in the workpiece surface. The foregoing composite rule also seemed to be applicable for steel SAE 1112 in the as-received condition.

3 The single tool force inferred from intercept tool forces, named tool-edge force, as well as possible rubbing on the flank, made the results obtained with alloy SAE 4135 as-received in flank-area studies less lucid. However, the explanation that there is rubbing and therefore friction at the flank at small clearance angles appears to be confirmed by some of the experimental data.

Acknowledgments

The authors wish to express their thanks to the Machinability Section, located at the Rock Island Arsenal, without whose support this investigation would have been difficult to carry out. They also wish to thank the University of California and the Department of Industrial Engineering for making available space and equipment. They further wish to thank Mr. Petersen, Senior Mechanician in the Industrial Engineering shops, for his able advice and many helpful suggestions during the life of the project. Special thanks also are given to Mr. Kot, who prepared all final curves, and Mrs. Arneson for typing the manuscript. Last but not least, they wish to express their sincerest appreciation to each of the engineering students at the University of California who worked on this project in the form of preparing specimens, testing, computation, analyzing, in fact, in every phase of work involved in this project. These students are: Eggleston, Hall, Harris, Herzog, MacDonald, Miller, Pabor, Saylor, and Thometz.

References

1 H. Takeyama and E. Usui, "The Effect of Tool-Chip Contact Area in Metal Machining," TRANS. ASME, vol. 80, 1958, pp. 1089–1096.

2 B. T. Chao and K. J. Trigger, "Controlled Contact Cutting Tools," TRANS. ASME, Series B, JOURNAL OF ENGINEERING FOR INDUSTRY, vol. 81, 1959, pp. 139–151.

3 S. Kobayashi, R. Herzog, D. M. Eggleston, and E. G. Thomsen, "An Application of Plastic Flow Analysis to Orthogonal Metal Cutting," University of California, IER Report No. 117, Issue No. 1, May, 1958, pp. 1–123.

4 D. R. Olberts, "A Study of the Effects of Tool Flank Wear on Tool-Chip Interface Temperature," TRANS. ASME, Series B, JOURNAL OF ENGINEERING FOR INDUSTRY, vol. 81, 1959, pp. 152–158.

5 M. Okoshi and T. Sata, "Friction on Relief Face of Cutting Tool," Scientific Papers of the Institute of Physical and Chemical Research, vol. 52, No. 1493, pp. 216–223, Tokyo, Japan, 1958.

6 S. Kobayashi and E. G. Thomsen, "An Application of Plastic Flow Analysis to Orthogonal Metal Cutting," University of California, IER Report No. 117, Issue No. 2, May, 1959, pp. 1–166.

7 S. Kobayashi, R. P. Herzog, D. M. Eggleston, and E. G. Thomsen, "A Critical Comparison of Metal-Cutting Theories With New Experimental Data," published in this issue, pp. 333–347.

8 I. Finnie and M. C. Shaw, "The Friction Process in Metal-Cutting," TRANS. ASME, vol. 78, 1956, pp. 1649–1657.

9 R. Holm, "Electric Contacts," H. Geber, Stockholm, Sweden, 1946, p. 214, also, "Mechanical Wear," TRANS. ASME, vol. 42, 1950, p. 317.

10 H. Ernst and M. E. Merchant, "Surface Friction of Clean Metals," Proceedings of Special Summer Conference on Friction and Surface Finish, Massachusetts Institute of Technology, Cambridge, Mass., 1940, p. 76.

11 F. P. Bowden and D. Tabor, "The Friction and Lubrication of Solids," Oxford University Press, London, England, 1954.

12 A. P. Green, "Friction Between Unlubricated Metals: A Theoretical Analysis of the Junction Model," *Proceedings of the Royal Society of London*, series A, vol. 228, 1955, pp. 191–204.

13 M. C. Shaw, P. A. Smith, N. H. Cook, and E. G. Loewen, "Leaded Steel and the Real Area of Contact in Metal-Cutting," TRANS. ASME, vol. 79, 1957, pp. 1143–1154.

14 M. C. Shaw and I. Finnie, "The Shear Stress in Metal-Cutting," TRANS. ASME, vol. 77, 1955, pp. 115–125.

15 A. W. J. Chisholm, "Progress Report on the Wear of Cutting Tools," Mechanical Engineering Research Laboratory, Glasgow, England, Plasticity Report 106, May, 1955.

S. KOBAYASHI[2]
Assistant Professor, Doshisha University, Kyoto, Japan

R. P. HERZOG
Design Engineer, Aerojet-General Corporation, Azusa, Calif.

D. M. EGGLESTON
Graduate Student in Mechanical Engineering, University of California, Berkeley, Calif.

E. G. THOMSEN
Professor of Metal Processing, University of California, Berkeley, Calif. Mem. ASME

A Critical Comparison of Metal-Cutting Theories With New Experimental Data[1]

The following alloys were investigated in orthogonal metal-cutting tests at varying cutting speeds, rake angles, and feeds: Steel SAE 1112 annealed and as-received; steel SAE 1020 as-received; steel SAE 4135, annealed, as-received, and Rockwell hardnesses 27 and 35–37; aluminum alloy 2024-T4; aluminum alloy 6061 in the O and T6 conditions; alpha brass. The new data obtained in the present investigation under wide variation of test conditions (speed for SAE 1112 annealed and as-received was varied from 0.083–1010 fpm) confirmed earlier observations [1, 3][3] that the shearing stresses are independent of the test conditions investigated. These observations were supported by theoretical considerations. Correlation of metal-cutting data with compression data on the basis of the incremental or distortion-energy theory was good for the ductile metals. Less perfect correlation was observed with SAE steel 4135 and aluminum alloy 2024-T4. It was found further that none of the known theoretical angle relationships were in agreement with all of the experimental data.

In spite of the considerable effort that research workers have expended in the past few years on metal-cutting research, the mechanics of this process are not fully understood. This is clearly a block in the road to improvement of the technology of conventional metal-cutting processes or to the development of new methods for materials which are difficult to machine.

Hence, the present studies were undertaken for the purpose of shedding additional light on the understanding of the mechanics of metal cutting. In order to achieve this objective, it was thought advisable to concentrate the principal effort on the tests given here which were designed to augment previous findings [1, 2, 3] and to reveal the following:

1 Effect of speed and structure on the shear stress in metal cutting.

2 Correlation of shear properties in metal cutting with those obtained from static tests.

3 Possible correlation of angle relationships with present theories.

Experimental Procedure

The tool materials employed in these orthogonal cutting tests consisted of only two types, in order to eliminate as much as possible the effect of tool material on the cutting conditions. These materials were HSS 18-4-1 (Rex AA) high-speed-steel tools for the low speeds and sintered-carbide (K3H) tools for the higher speeds. All tools had $^1/_8$-in. square shanks and were reground on their flanks on a universal tool grinder between each test run.

The grinding wheels were cup-shaped and were selected with alundum and diamond abrasives for refinishing the high-speed-steel and sintered-carbide tools, respectively. In addition, the faces of these tools were also refinished between each test run by orienting the lay in the direction of chip flow to a roughness of 5 microin. or better. This was accomplished by honing the faces of the high-speed steels with a hard synthetic Arkansas honing stone and those of the brazed sintered-carbide-tipped tools by lapping with a diamond abrasive (5000 mesh) and kerosene on a cast-iron lap. Each test run was of relatively short duration in order to reduce the effect of wear on the geometry of the cutting edge, but long enough to yield steady-state conditions. The length of these runs was of the order of from 10 to 20 revolutions of the workpiece.

All chip ratios at high cutting speeds were calculated from weight-length measurements of selected chips. The previous method of calculating chip ratios by making length measurements of continuous chips and dividing these lengths into the undeformed original lengths when still attached to the workpiece no longer was suitable. This was due to the fact that it was inadvisable to provide the tubular test specimen with a longitudinal reference slot because of the propensity of the cutting edge of the sintered-carbide tools to break out. It was for this reason that a chip-cutting device was constructed for preparing fixed known lengths of chips in order to shorten the time required to perform weight-length determinations of chip ratios. This punch-and-die set was arranged in such a way that chip segments in the form of wedges were cut out which approximately fitted the tube curvature from which they were removed. It was, therefore, only necessary to collect 40 cut chips (a number arbitrarily selected for all measurements) and to weigh these chips on a sensitive balance. The total center-line distance of 40 chips, with a die spacing of 0.2375 in., was 9.50 in.

The low-speed tests which were made with SAE 1112 steels in the annealed and as-received state required a special drive and measuring setup. The American Pacemaker lathe, on which all tests were made, was provided with an auxiliary d-c drive, which permitted obtaining constant surface speeds as low as 0.083 fpm. Some difficulties were encountered in obtaining continuous feeds at the low spindle speeds, since the carriage exhibited a stick-slip action. This caused the force readings to fluctuate and the feed to become unsteady. After considerable

---

[1] This paper is taken in part from the report, "An Application of Plastic Flow Analysis to Orthogonal Metal Cutting," Institute of Engineering Research, University of California, Series No. 117, Issue No. 2, May, 1959. The investigation was sponsored by the Machinability Program located at Rock Island Arsenal under Contract DA-11-070-508-ORD-605 with the University of California.

[2] At present, Graduate Student in Mechanical Engineering, University of California, Berkeley, Cal.

[3] Numbers in brackets designate references at end of paper.

Contributed by the Metal Processing Activity of the Production Engineering Division and presented at the Annual Meeting, Atlantic City, N. J., November 29–December 4, 1959, of The American Society of Mechanical Engineers.

Note: Statements and opinions advanced in papers are to be understood as individual expressions of their authors and not those of the Society. Manuscript received at ASME Headquarters, June 16, 1959. Paper No. 59—A-131.

experimentation in attempting to eliminate this stick-slip action of the carriage, it was concluded that it was not possible without extensive modifications of the lathe. It was, therefore, decided to provide an event-marking system, by which the instantaneous chip ratios could be correlated with the corresponding dynamometer-force readings. This was accomplished by providing the test specimens with a number of shallow external grooves, which actuated a microswitch through a ball rider, without apparently affecting the metal-cutting action. A small applied voltage then gave a series of unequally spaced marks on the Sanborn recording tape together with the dynamometer-force readings, so that these events could be correlated. In addition, actual instantaneous carriage motions and tool-dynamometer deflections were used in order to determine the actual feeds. The chip ratios were then obtained by dividing the chip thicknesses by the instantaneous feeds.

The static properties of the test materials were obtained by compression tests of small specimens. The method consisted of preparing cylindrical specimens from the walls of the test specimens keeping the diameter-to-length ratio near one. The ends were provided with shallow spiral grooves in order to aid in lubrication. The compression tests were carried out in a die set having parallel polished anvils, which was mounted between head and platen of a hydraulic testing machine. In addition, microtension tests were performed for the four alloy steels, SAE-4135 metallurgical structures, steel SAE 1020, and aluminum alloy 2024-T4, in order to compare the tension properties with compression properties. The tension test specimens had an initial diameter of approximately $^1/_8$ in. All stresses were corrected for the combined stress effect after necking, by measuring the neck radius and applying the Bridgman [4] correction.

The results of the present tests (these are in addition to those published earlier [1, 2, 3]) are given in Tables 1 to 5, inclusive. For reasons of economizing on space, the low-speed data obtained with the free-cutting steels are not given in tabular form.

Discussion of Results

Shear Stresses. The resolved shearing forces F_s for the alloys and test conditions reported earlier [1, 3] were found to be linear functions of the shear areas A_s on which they acted. Fig. 1, by way of example, shows the results for all test conditions for steel SAE 1112 annealed and steel SAE 1112 as received, respectively. The intercept was always of finite magnitude and positive at zero A_s. This was interpreted to mean that this intrecept force was not associated with shearing on the shear plane and contributed nothing to chip deformation. In the light of recent experiments, however, it seems (at least in part) that the intercept may be a singular force at the tool edge. This assumption was also made by Takeyama and Usui [5] and more recently by Vieregge [6]. The latter author likened the local force per unit area to a stagnation pressure at the edge of the tool, not unlike that occurring in a stream flowing past an obstacle. Assuming now that the intercept does not contribute to chip deformation on the shear plane, then the average shearing stress is given by,

$$\tau = \frac{F_s - F_{s0}}{A_s} \tag{1}$$

where F_{s0} is the shear force when $A_s = 0$, and τ is simply the slope of the straight line. It follows from this that the shearing

Symbols used in Tables 1–5

α = rake angle, deg
F_H = force component in direction of motion, lb
F_V = force component normal to direction of motion, lb
t_0 = feed, 1/1000 in. per revolution
r_c = chip ratio
w_0 = wall thickness = initial width of chip, in.
V = cutting speed, fpm
χ = flank clearance angle, deg
ϕ = shear angle, deg

Table 1 Cutting data

Work Material - Steel SAE 1112 As-Received Cutting Fluid - Air

Tool Material - HSS 18-4-1 (REX AA) (through 10 fpm) Outside Diameter of Tube 2.875"
Sintered Carbide (K3H)

Run No.	α	F_H	F_V	t_0	r_c	w_0	χ	V	ϕ	Run No.	α	F_H	F_V	t_0	r_c	w_0	χ	V	ϕ
741	25	84	10	1.99	0.480	0.201	3	10.0	28.6	972	20	340	67	9.81	0.573	0.200	1	384	33.9
742	25	186	17	4.89	0.448	0.201	3	10.0	26.6	771	15	113	30	2.75	0.555	0.200	1	453	32.1
743	25	261	20	7.35	0.436	0.201	3	10.0	25.9	767	15	187	48	4.89	0.558	0.200	1	453	32.3
744	25	339	24	9.81	0.427	0.201	3	10.0	25.4	768	15	190	47	4.89	0.556	0.200	1	453	32.1
745	40	69	- 2	1.99	0.607	0.201	3	10.0	37.3	769	15	290	96	7.35	0.524	0.200	1	453	30.4
746	40	142	-16	4.89	0.596	0.201	3	10.0	36.5	770	15	396	141	9.81	0.508	0.200	1	453	29.5
747	40	198	-26	7.35	0.595	0.201	3	10.0	36.2	985	0	199	158	2.75	0.427	0.200	1	1010	23.1
748	40	256	-34	9.81	0.584	0.201	3	10.0	35.6	988	0	205	176	2.75	0.451	0.200	1	1010	24.3
976	0	137	68	2.75	0.450	0.200	1	384	24.2	1259	0	195	156	2.75	0.433	0.200	1	1010	23.4
980	0	142	63	2.75	0.445	0.200	1	384	24.0	984	0	290	190	4.89	0.469	0.200	1	1010	25.1
977	0	258	148	4.89	0.422	0.200	1	384	22.9	989	0	287	190	4.89	0.474	0.200	1	1010	25.4
981	0	255	145	4.89	0.428	0.200	1	384	23.2	991	0	287	190	4.89	0.477	0.200	1	1010	25.5
978	0	380	225	7.35	0.417	0.200	1	384	22.6	1260	0	283	186	4.89	0.483	0.200	1	1010	25.8
982	0	380	217	7.35	0.432	0.200	1	384	23.4	986	0	364	188	7.35	0.528	0.200	1	1010	27.8
979	0	480	273	9.81	0.445	0.200	1	384	24.0	990	0	357	173	7.35	0.515	0.200	1	1010	27.2
983	0	483	278	9.81	0.446	0.200	1	384	24.0	1261	0	350	167	7.35	0.526	0.200	1	1010	27.7
965	20	105	23	2.75	0.578	0.200	1	384	34.1	1263	0	350	167	7.35	0.518	0.200	1	1010	27.4
969	20	105	21	2.75	0.568	0.200	1	384	33.5	987	0	440	185	9.81	0.572	0.200	1	1010	29.8
966	20	169	26	4.89	0.597	0.200	1	384	35.1	992	0	430	183	9.81	0.557	0.200	1	1010	29.1
970	20	172	25	4.89	0.583	0.200	1	384	34.4	1262	0	438	184	9.81	0.556	0.200	1	1010	29.1
967	20	254	40	7.35	0.593	0.200	1	384	35.0	776	25	122	39	2.75	0.572	0.200	1	1010	34.3
971	20	249	38	7.35	0.594	0.200	1	384	35.0	775	25	194	50	4.89	0.567	0.200	1	1010	34.0
968	20	342	67	9.81	0.580	0.200	1	384	34.2	774	25	262	46	7.35	0.605	0.200	1	1010	36.4

Table 2 Cutting data

Work Material - Steel SAE 1020 As-Received Cutting Fluid - Air

Tool Material - Sintered Carbide (K3H) Outside Diameter of Tube - 2.875"

Run No.	α	F_H	F_v	t_o	r_c	w_o	χ	V	ϕ	Run No.	α	F_H	F_v	t_o	r_c	w_o	χ	V	ϕ
949	0	250	250	2.75	0.227	0.200	1	384	12.8	964	20	443	230	9.81	0.490	0.200	1	384	29.0
953	0	250	254	2.75	0.239	0.200	1	384	13.4	1251	0	230	237	2.75	0.272	0.201	1	530	15.2
1255	0	255	260	2.75	0.227	0.201	1	384	12.8	1241	0	330	295	4.89	0.331	0.201	3	530	18.3
950	0	355	328	4.89	0.294	0.200	1	384	16.4	1244	0	308	280	4.89	0.381	0.201	3	530	20.9
954	0	360	335	4.89	0.284	0.200	1	384	15.9	1248	0	340	312	4.89	0.322	0.201	1	530	17.8
1256	0	360	330	4.89	0.281	0.201	1	384	15.7	1252	0	355	328	4.89	0.316	0.201	1	530	17.5
951	0	476	426	7.35	0.326	0.200	1	384	18.1	1242	0	410	330	7.35	0.426	0.201	3	530	23.1
955	0	475	419	7.35	0.336	0.200	1	384	18.6	1245	0	420	340	7.35	0.426	0.201	3	530	23.1
1257	0	485	425	7.35	0.315	0.201	1	384	17.5	1253	0	445	352	7.35	0.386	0.201	1	530	21.1
952	0	570	460	9.81	0.381	0.200	1	384	20.9	1243	0	510	350	9.81	0.458	0.201	3	530	24.6
956	0	570	460	9.81	0.383	0.200	1	384	21.0	1246	0	505	355	9.81	0.452	0.201	3	530	24.3
1258	0	565	455	9.81	0.381	0.201	1	384	20.9	1254	0	540	395	9.81	0.453	0.201	1	530	24.4
957	20	184	135	2.75	0.308	0.200	1	384	17.9	1196	0	217	204	2.75	0.337	0.200	1	1010	18.6
957A	20	147	83	2.75	0.436	0.200	1	384	25.7	1200	0	230	216	2.75	0.336	0.200	1	1010	18.6
961	20	145	85	2.75	0.427	0.200	1	384	25.1	1197	0	294	228	4.89	0.447	0.200	1	1010	24.1
958	20	270	169	4.89	0.403	0.200	1	384	23.7	1201	0	313	252	4.89	0.432	0.200	1	1010	23.4
962	20	233	126	4.89	0.484	0.200	1	384	28.6	1198	0	361	234	7.35	0.513	0.200	1	1010	27.2
959	20	363	205	7.35	0.454	0.200	1	384	26.6	1202	0	369	236	7.35	0.502	0.200	1	1010	26.7
963	20	363	204	7.35	0.450	0.200	1	384	26.6	1199	0	440	272	9.81	0.585	0.200	1	1010	30.3
960	20	443	231	9.81	0.502	0.200	1	384	29.7	1203	0	462	270	9.81	0.554	0.200	1	1010	29.0

Table 3 Cutting data

Work Material - Alloy Steel SAE 4135 Annealed (RB-81) Cutting Fluid - Air

Tool Material - Sintered Carbide (K3H) Outside Diameter of Tube - 3.947"

Run No.	α	F_H	F_v	t_o	r_c	w_o	χ	V	ϕ	Run No.	α	F_H	F_v	t_o	r_c	w_o	χ	V	ϕ
857	0	300	306	2.75	0.221	0.200	1	334	12.5	875	0	510	300	7.35	0.420	0.200	1	540	22.8
858	0	410	342	4.89	0.290	0.200	1	334	16.2	877	0	620	320	9.81	0.456	0.200	1	540	24.5
861	0	538	385	7.35	0.341	0.200	1	334	18.8	793	0	238	205	2.75	0.301	0.200	3	540	16.8
862	0	645	412	9.81	0.405	0.200	1	334	22.0	795	0	345	243	4.89	0.386	0.200	3	540	21.1
865	0	415	355	4.89	0.288	0.200	1	334	16.1	797	0	482	290	7.35	0.417	0.200	3	540	22.6
867	0	530	377	7.35	0.356	0.200	1	334	19.6	799	0	595	323	9.81	0.444	0.200	3	540	24.0
869	0	655	420	9.81	0.379	0.200	1	334	20.7	872	0	240	215	2.75	0.292	0.200	3	540	16.3
860	0	400	345	4.89	0.304	0.200	3	334	16.9	874	0	370	258	4.89	0.364	0.200	3	540	20.0
863	0	524	367	7.35	0.350	0.200	3	334	19.3	876	0	480	275	7.35	0.424	0.200	3	540	23.0
864	0	635	402	9.81	0.392	0.200	3	334	21.4	878	0	600	295	9.81	0.459	0.200	3	540	24.7
866	0	405	330	4.89	0.293	0.200	3	334	16.3	929	20	186	100	2.75	0.375	0.200	1	393	22.0
868	0	530	380	7.35	0.343	0.200	3	334	18.9	930	20	342	210	4.89	0.334	0.200	1	393	19.5
870	0	640	400	9.81	0.393	0.200	3	334	21.4	931	20	429	206	7.35	0.400	0.200	1	393	23.6
792	0	263	243	2.75	0.338	0.200	1	540	18.7	932	20	550	225	9.81	0.422	0.200	1	393	24.8
794	0	390	295	4.89	0.380	0.200	1	540	20.8	933	20	193	104	2.75	0.374	0.200	1	393	22.0
796	0	513	323	7.35	0.424	0.200	1	540	23.0	934	20	343	210	4.89	0.329	0.200	1	393	19.2
798	0	610	346	9.81	0.442	0.200	1	540	23.8	935	20	443	226	7.35	0.387	0.200	1	393	22.7
871	0	271	240	2.75	0.300	0.200	1	540	16.7	936	20	552	236	9.81	0.422	0.200	1	393	24.8
873	0	400	285	4.89	0.363	0.200	1	540	20.0										

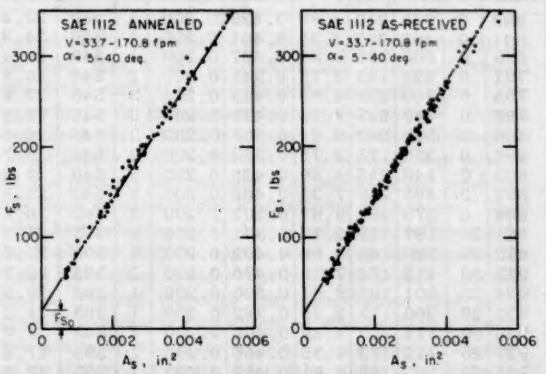


Fig. 1 Shear force as function of shear-plane area for steel SAE 1112 references [1] and [3]

stresses for the test conditions investigated appear to be invariant.

In consequence of the foregoing, it seemed of interest to expand the range of cutting conditions in order to study further the behavior of the shearing stresses. This was done and the results of the experiments are shown in Figs. 2 to 4, inclusive. For all alloys for which straight lines were previously established [1, 3], identical straight lines with the same slopes and intercepts were reproduced in these figures. For the additional alloys previously not tested, such as steel SAE 1020, Fig. 2, alloy steel SAE 4135 in the four metallurgical states, Fig. 3, and aluminum alloy 6061-0, Fig. 4, additional straight lines were drawn, which represent the best curves through the experimental points. The magnitude of all slopes have been determined and are compiled in Table 6 as shear stresses.

It appears worth while to appraise further the theoretical possibility of an invariant shear stress. This may be done profitably by examining possible theoretical strain rates. If it is assumed, for example, that the forward rate of heat flow is unable to alter

Table 4 Cutting data

Work Material - Aluminum Alloy 6061-0 Cutting Fluid - Air

Tool Material - Sintered Carbide (K3H) Outside Diameter of Tube - 4.150"

Run No.	α	F_H	F_v	t_o	r_c	w_o	χ	V	ϕ	Run No.	α	F_H	F_v	t_o	r_c	w_o	χ	V	ϕ
1009	20	50	33	1.99	0.258	0.200	1	2080	14.8	1012	20	129	63	7.35	0.324	0.200	1	2080	18.9
1014A	20	48	34	1.99	0.252	0.200	1	2080	14.6	1017	20	130	58	7.35	0.321	0.200	1	2080	18.8
1014B	20	55	42	1.99	0.205	0.200	1	2080	11.7	1013	20	160	80	9.81	0.339	0.200	1	2080	19.9
1014C	20	65	56	1.99	0.170	0.200	1	2080	9.7	1018	20	155	68	9.81	0.353	0.200	1	2080	20.7
1010	20	65	38	2.75	0.274	0.200	1	2080	15.9	1019	20	63	52	1.99	0.175	0.200	1	1280	10.0
1015	20	62	37	2.75	0.270	0.200	1	2080	15.6	1020	20	78	56	2.75	0.194	0.200	1	1280	11.1
1011	20	94	47	4.89	0.318	0.200	1	2080	18.6	1021	20	103	58	4.89	0.258	0.200	1	1280	14.9
1016	20	95	48	4.89	0.314	0.200	1	2080	18.3	1022	20	140	77	7.35	0.259	0.200	1	1280	15.0
										1023	20	175	85	9.81	0.296	0.200	1	1280	17.2

Work Material - Aluminum Alloy 2024-T4 Cutting Fluid - Air

Tool Material - Sintered Carbide (K3H) Outside Diameter of Tube - 4.168"

Run No.	α	F_H	F_v	t_o	r_c	w_o	χ	V	ϕ	Run No.	α	F_H	F_v	t_o	r_c	w_o	χ	V	ϕ
997	0	72	35	1.99	0.565	0.200	1	2080	29.5	1007	20	77	18	2.75	0.606	0.200	1	2080	35.7
998	0	73	34	1.99	0.575	0.200	1	2080	29.9	1008	20	78	16	2.75	0.602	0.200	1	2080	35.5
994	0	90	36	2.75	0.618	0.200	1	2080	31.7	1001	20	119	11	4.89	0.659	0.200	1	2080	38.6
995	0	90	37	2.75	0.616	0.200	1	2080	31.6	1002	20	118	12	4.89	0.655	0.200	1	2080	38.4
993	0	139	42	4.89	0.765	0.200	1	2080	37.4	1003	20	162	0	7.35	0.737	0.200	1	2080	42.7
996	0	139	43	4.89	0.778	0.200	1	2080	37.9	1004	20	157	1	7.35	0.711	0.200	1	2080	41.4
999	20	62	18	1.99	0.571	0.200	1	2080	33.7	1006	20	205	-10	9.81	0.810	0.200	1	2080	46.2
1000	20	62	18	1.99	0.583	0.200	1	2080	34.4										

Work Material - Aluminum Alloy 6061-T6 Cutting Fluid - Air

Tool Material - Sintered Carbide (K3H) Outside Diameter of Tube - 4.150"

Run No.	α	F_H	F_v	t_o	r_c	w_o	χ	V	ϕ	Run No.	α	F_H	F_v	t_o	r_c	w_o	χ	V	ϕ
1212	20	69	28	2.75	0.480	0.200	1	2070	28.4	1214	20	138	32	7.35	0.593	0.200	1	2070	35.0
1216	20	69	27	2.75	0.470	0.200	1	2070	27.8	1218	20	138	30	7.35	0.582	0.200	1	2070	34.3
1213	20	96	33	4.89	0.538	0.200	1	2070	31.8	1215	20	170	28	9.81	0.620	0.200	1	2070	36.5
1217	20	102	30	4.89	0.538	0.200	1	2070	31.8	1219	20	178	28	9.81	0.617	0.200	1	2070	36.3

Work Material - Alloy Steel SAE 4135 As-Received (RC-18) Cutting Fluid - Air

Tool Material - Sintered Carbide (K3H) Outside Diameter of Tube - 3.946"

Run No.	α	F_H	F_v	t_o	r_c	w_o	χ	V	ϕ	Run No.	α	F_H	F_v	t_o	r_c	w_o	χ	V	ϕ
879	0	300	275	2.75	0.245	0.200	1	334	13.8	889	0	378	240	4.89	0.429	0.200	1	540	23.2
881	0	410	307	4.89	0.340	0.200	1	334	18.8	891	0	495	285	7.35	0.461	0.200	1	540	24.8
883	0	530	356	7.35	0.400	0.200	1	334	21.8	893	0	600	305	9.81	0.497	0.200	1	540	26.4
885	0	645	380	9.81	0.432	0.200	1	334	23.4	791	0	228	183	2.75	0.341	0.200	3	540	18.8
895	0	275	250	2.75	0.261	0.200	1	334	14.6	785	0	340	228	4.89	0.423	0.200	3	540	22.9
897	0	407	307	4.89	0.328	0.200	1	334	18.2	787	0	460	257	7.35	0.477	0.200	3	540	25.5
899	0	530	347	7.35	0.394	0.200	1	334	21.5	789	0	567	282	9.81	0.505	0.200	3	540	26.8
901	0	636	385	9.81	0.432	0.200	1	334	23.4	888	0	230	175	2.75	0.338	0.200	3	540	18.7
882	0	395	285	4.89	0.339	0.200	3	334	18.7	890	0	340	215	4.89	0.420	0.200	3	540	22.8
884	0	513	330	7.35	0.388	0.200	3	334	21.2	892	0	465	250	7.35	0.468	0.200	3	540	25.1
886	0	620	370	9.81	0.438	0.200	3	334	23.7	894	0	575	290	9.81	0.502	0.200	3	540	26.7
896	0	275	250	2.75	0.255	0.200	3	334	14.3	921	20	197	113	2.75	0.371	0.200	1	393	21.7
898	0	392	288	4.89	0.333	0.200	3	334	18.4	922	20	320	163	4.89	0.402	0.200	1	393	23.7
900	0	507	328	7.35	0.392	0.200	3	334	21.4	923	20	412	174	7.35	0.475	0.200	1	393	28.1
902	0	606	358	9.81	0.442	0.200	3	334	23.8	924	20	501	165	9.81	0.500	0.200	1	393	29.5
790	0	260	228	2.75	0.362	0.200	1	540	19.9	925	20	200	113	2.75	0.362	0.200	1	393	21.2
784	0	370	252	4.89	0.445	0.200	1	540	24.0	926	20	314	160	4.89	0.406	0.200	1	393	23.9
786	0	480	283	7.35	0.491	0.200	1	540	26.2	927	20	412	173	7.35	0.466	0.200	1	393	27.6
788	0	582	307	9.81	0.512	0.200	1	540	27.1	928	20	500	163	9.81	0.498	0.200	1	393	29.5
887	0	257	210	2.75	0.358	0.200	1	540	19.7										

Table 4 (continued)

Work Material - Alloy Steel SAE 4135 (RC-26) Cutting Fluid - Air

Tool Material - Sintered Carbide (K3H) Outside Diameter of Tube - 3.950"

Run No.	α	F_H	F_v	t_o	r_c	w_o	χ	V	ϕ	Run No.	α	F_H	F_v	t_o	r_c	w_o	χ	V	ϕ
841	0	272	222	2.75	0.408	0.200	1	334	22.2	801	0	244	195	2.75	0.545	0.201	3	540	28.6
849	0	268	218	2.75	0.415	0.200	1	334	22.5	803	0	360	225	4.89	0.586	0.201	3	540	30.4
842	0	390	255	4.89	0.530	0.200	1	334	27.9	809	0	350	215	4.89	0.613	0.201	3	540	31.5
850	0	382	245	4.89	0.531	0.200	1	334	28.0	805	0	482	233	7.35	0.661	0.201	3	540	33.6
845	0	515	273	7.35	0.594	0.200	1	334	30.7	807	0	590	220	9.81	0.678	0.201	3	540	34.1
853	0	515	272	7.35	0.589	0.200	1	334	30.5	808	0	590	230	9.81	0.666	0.200	3	540	33.7
846	0	640	320	9.81	0.650	0.200	1	334	33.0	941	0	238	174	2.75	0.519	0.200	3	540	27.4
854	0	630	295	9.81	0.650	0.200	1	334	33.0	942	0	370	225	4.89	0.612	0.200	3	540	31.5
843	0	250	193	2.75	0.417	0.200	3	334	22.6	943	0	465	208	7.35	0.640	0.200	3	540	32.6
851	0	258	203	2.75	0.411	0.200	3	334	22.3	944	0	595	250	9.81	0.660	0.200	3	540	33.4
844	0	356	210	4.89	0.515	0.200	3	334	27.2	945	0	244	186	2.75	0.528	0.200	3	540	27.8
852	0	370	220	4.89	0.511	0.200	3	334	27.1	946	0	351	192	4.89	0.603	0.200	3	540	31.6
847	0	497	253	7.35	0.594	0.200	3	334	30.7	947	0	476	206	7.35	0.636	0.200	3	540	32.5
855	0	478	238	7.35	0.611	0.200	3	334	31.4	948	0	587	226	9.81	0.667	0.200	3	540	33.7
848	0	605	261	9.81	0.665	0.200	3	334	33.6	911	20	236	140	2.75	0.446	0.200	1	393	26.3
856	0	625	270	9.81	0.655	0.200	3	334	33.2	915	20	233	148	2.75	0.466	0.200	1	393	27.6
800	0	266	237	2.75	0.550	0.201	1	540	28.8	919	20	235	136	2.75	0.450	0.200	1	393	26.6
802	0	378	250	4.89	0.642	0.201	1	540	32.7	912	20	325	133	4.89	0.535	0.200	1	393	31.7
804	0	497	262	7.35	0.640	0.201	1	540	32.6	916	20	317	126	4.89	0.563	0.200	1	393	33.3
806	0	615	275	9.81	0.885	0.201	1	540	41.5	920	20	307	120	4.89	0.547	0.200	1	393	32.3
937	0	258	205	2.75	0.545	0.200	1	540	28.6	913	20	402	108	7.35	0.622	0.200	1	393	36.6
938	0	378	226	4.89	0.612	0.200	1	540	31.5	917	20	412	106	7.35	0.638	0.200	1	393	37.5
939	0	495	237	7.35	0.644	0.200	1	540	32.8	914	20	508	93	9.81	0.655	0.200	1	393	38.4
940	0	598	251	9.81	0.686	0.200	1	540	34.5	918	20	515	93	9.81	0.654	0.200	1	393	38.4

Work Material - Alloy Steel SAE 4135 (RC-35~37) Cutting Fluid - Air

Tool Material - Sintered Carbide (K3H) Outside Diameter of Tube - 3.942"

Run No.	α	F_H	F_v	t_o	r_c	w_o	χ	V	ϕ	Run No.	α	F_H	F_v	t_o	r_c	w_o	χ	V	ϕ
826	0	268	188	2.75	0.618	0.200	1	334	31.7	831	0	647	257	9.81	0.771	0.200	3	334	37.6
832	0	255	185	2.75	0.622	0.200	1	334	31.9	836	0	250	162	2.75	0.700	0.200	1	540	35.0
823	0	395	225	4.89	0.743	0.200	1	334	36.6	837	0	250	175	2.75	0.709	0.200	1	540	35.3
833	0	390	223	4.89	0.728	0.200	1	334	36.1	838	0	360	187	4.89	0.754	0.200	1	540	37.0
821	0	525	254	7.35	0.760	0.200	1	334	37.2	840	0	360	195	4.89	0.756	0.200	1	540	37.1
834	0	523	242	7.35	0.755	0.200	1	334	37.1	811	0	247	177	2.75	0.688	0.200	3	540	34.5
819A	0	655	285	9.81	0.783	0.200	1	334	38.1	813	0	375	203	4.89	0.754	0.200	3	540	37.0
835	0	660	284	9.81	0.760	0.200	1	334	37.2	833A	0	360	187	4.89	0.772	0.200	3	540	37.7
825	0	255	169	2.75	0.606	0.200	3	334	31.2	815	0	490	243	7.35	0.796	0.200	3	540	38.6
827	0	250	180	2.75	0.617	0.200	3	334	31.7	903	20	228	101	2.75	0.623	0.200	1	393	33.5
824	0	378	204	4.89	0.724	0.200	3	334	35.9	907	20	227	98	2.75	0.628	0.200	1	393	36.9
828	0	380	202	4.89	0.716	0.200	3	334	35.5	904	20	313	77	4.89	0.732	0.200	1	393	42.5
822	0	510	227	7.35	0.789	0.200	3	334	38.2	908	20	320	82	4.89	0.732	0.200	1	393	42.5
829	0	515	230	7.35	0.753	0.200	3	334	37.0	905	20	428	88	7.35	0.800	0.200	1	393	46.1
830	0	515	227	7.35	0.760	0.200	3	334	37.2	909	20	410	56	7.35	0.807	0.200	1	393	46.4
820	0	640	244	9.81	0.764	0.200	3	334	37.3										

significantly the temperature of the metal flowing toward the shear zone, then it may be assumed that plastic deformation begins with a metal at or near the workpiece temperature. This temperature in the present case was room temperature. It must be admitted, then, that it is impossible for thermal fluctuations to assist greatly in the movement of dislocations over barriers. This was precisely the argument that was advanced previously [1, 3] which appeared to give a satisfactory explanation. Good support for this is given by Cottrell [7] who would estimate under these conditions that the applied shear stress must exceed the internal shear stress in order that rapid plastic deformation be initiated. Cottrell visualizes the internal-stress field as an undulating one due to such effects as work-hardening, precipitation-hardening, and grain boundaries. It appears, therefore, that the initiation of plastic deformation under these isothermal conditions depends only on the metallurgical structure of the alloy.

It may next be asked if the expected possible isothermal strain rate $\dot{\epsilon}_p$ is high enough to exceed the actual or required shear-strain rate $\dot{\epsilon}_r$ in metal cutting; for if this were not the case, the possible shear-strain rate $\dot{\epsilon}_p$ would tend to lag behind the required rate $\dot{\epsilon}_r$. Since this condition from consideration of continuity cannot exist, it must be assumed that a higher applied stress is necessary to speed up the process. Thus an increasing stress would be required, whenever

$$\dot{\epsilon}_r > \dot{\epsilon}_p \qquad (2)$$

at the same internal state of stress and initial test condition. In order to test this relationship we may make comparisons with published data. Thus for metal cutting the strain rates $\dot{\epsilon}_r$ appear to be estimated to vary from about 10^3 to 10^5 per second [1]. In order to compare this with a possible strain rate $\dot{\epsilon}_p$ of the metal, we might again consult Cottrell [7]. This author

Table 5 Cutting data

Work Material - Alpha Brass Cutting Fluid - Air

Tool Material - Sintered Carbide (K3H) Outside Diameter of Tube - 3.486"

Run No.	α	F_H	F_v	t_o	r_c	w_o	χ	V	ϕ	Run No.	α	F_H	F_v	t_o	r_c	w_o	χ	V	ϕ
1204	20	97	30	2.75	0.361	0.183	1	1720	21.2	1206	20	240	72	7.35	0.397	0.183	1	1720	23.4
1208	20	94	28	2.75	0.384	0.183	1	1720	22.5	1210	20	230	71	7.35	0.404	0.183	1	1720	23.7
1205	20	165	52	4.89	0.392	0.183	1	1720	23.0	1207	20	305	87	9.81	0.406	0.183	1	1720	23.9
1209	20	160	48	4.89	0.395	0.183	1	1720	23.2	1211	20	302	84	9.81	0.408	0.183	1	1720	24.1

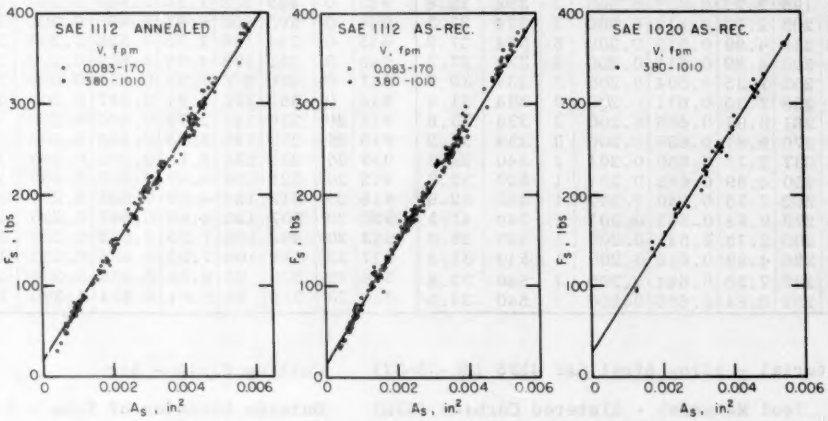


Fig. 2 Shear force as function of shear-plane area for steel SAE 1112 and steel SAE 1020. Data from Tables 1 and 2 and Fig. 1.

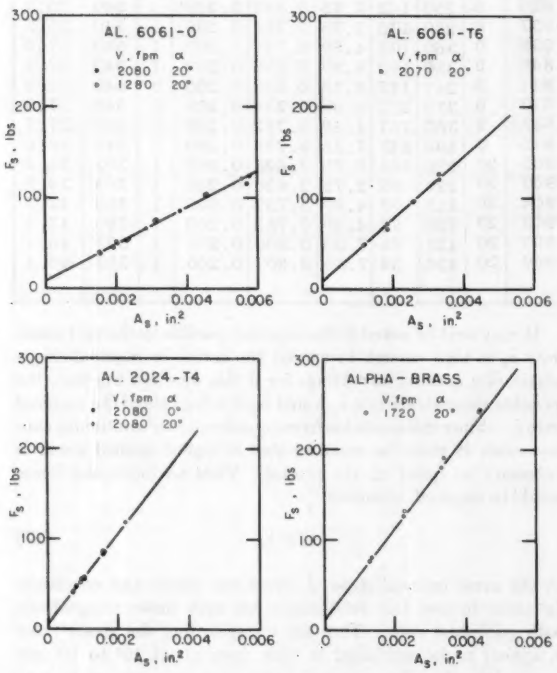


Fig. 3 Shear force as function of shear plane area for alloy steel SAE 4135. Data from Table 3.

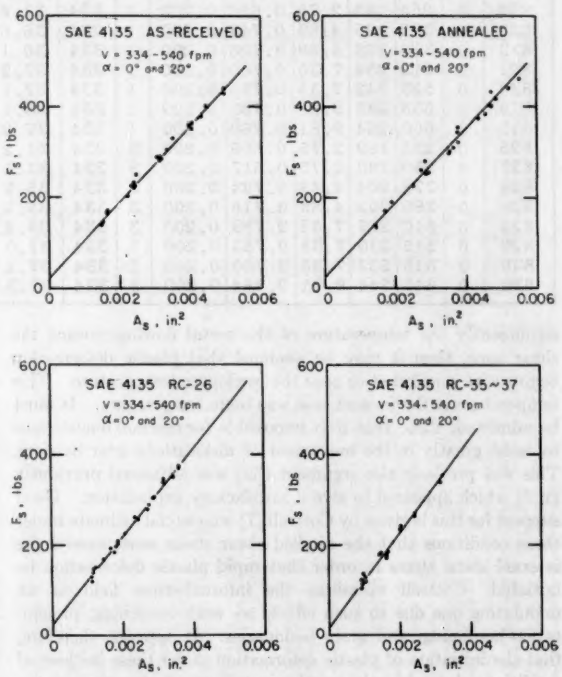


Fig. 4 Shear force as function of shear plane area for aluminum alloys and alpha-brass. Data from Tables 4 and 5.

Table 6 Shear stresses from metal-cutting tests

Material	Condition	Shear stress, psi	Speed range, fpm
Steel SAE 1112....	Annealed	60500	0.083–1010
Steel SAE 1112....	As-received (cold-finished)	63500	0.083–1010
Steel SAE 1020....	As-received (cold-finished)	61500	384–1010
Steel SAE 4135....	Annealed	81500	334–540
Steel SAE 4135....	As-received RC-18 (cold-finished)	84500	334–540
Steel SAE 4135....	RC-27 (heat-treated)	96700	334–540
Steel SAE 4135....	RC-35 to 37 (heat-treated)	102500	334–540
Alum 2024-T4....	Extruded	50000	2200
Alum 6061-0.....	Extruded	20500	1350–2200
Alum 6061-T6....	Extruded	34200	2200
Alpha-brass.....	As-received (cold-drawn)	50600	2000

estimates a strain rate for a so-called glissile dislocation by the following equation:

$$\dot{\epsilon}_p = \rho b v \tag{3}$$

where ρ is the number of dislocations per unit area moving at velocity v, and b is the unit of slip associated with the dislocation (usually one atomic spacing). The estimated values are

$$\rho \approx 10^8 \text{ to } 10^9 \text{ per sq cm}$$

$$v \approx 10^3 \text{ cm per sec}$$

$$b \approx 3 \times 10^{-8} \text{ cm}$$

hence $\dot{\epsilon}_p = 3 \times 10^3$ to 3×10^4 per sec. Whence it is apparent that $\dot{\epsilon}_p$ and $\dot{\epsilon}_r$ are of about the same order of magnitude and that a rapid strain rate without a further rise in stress is possible. Furthermore, Cottrell estimates that after a time of 10^{-9} to 10^{-8} sec, the rate of plastic strain increases rapidly with the multiplication of dislocations. It may be implied also from examining Equation (2) that supercutting speeds will affect the shear stresses only then when the inequality $\dot{\epsilon}_r > \dot{\epsilon}_p$ exists.

For the very low metal-cutting speeds employed in the present investigation, the deformation process also appears to be isothermal, at least when considering the bulk of the material. There appeared to be no noticeable rise in temperature in the metal-cutting zone, even though the strain rates were still of the order of possibly 10^2 per second or greater. Thus all heat energy was dissipated away from the shear zone and the bulk flow was isothermal. This does not preclude, however, that catastrophic plastic slip could have occurred (with a local rise in temperature) as has been observed by Basinski [8] in aluminum at very low test temperatures and low speeds, if one accepts the fact that deformation in metal cutting is intermittent. It appears from the fact, however, that the shear stresses in metal cutting at these low speeds (which are still of the order of impact speeds with ordinary test methods) remain constant, that no appreciable contribution from thermal fluctuations can be present in aiding dislocations to overcome barriers. Hence, at the low cutting speeds as well as those of intermediate magnitude, the internal stress field also appears to be the major factor in determining the flow condition. Thus for conventional metal-cutting speeds, the dynamic shear stress must be regarded as constant. A similar conclusion was reached by Chao and Trigger [9], who regard the dynamic shear stress as a fundamental property of the material.

Correlation With Static Properties. It was shown previously [1, 3] that the correlation of metal-cutting data with static properties was good. Hence, a similar comparison will be made in this section for the additional data now available. However, it will be of benefit at this point to review the theories of plastic deformation which make a comparison of results possible for data obtained with different test methods.

Apparently the most logical theory for relating plastic strains to stresses is the incremental theory, also referred to as the distortion-energy theory. This theory asserts that incremental plastic strains are proportional to deviator stresses. Hence, let the equations to follow be the accepted plasticity equations:

$$\left.\begin{aligned} d\epsilon_x &= {}^3/_2 \frac{d\bar{\epsilon}}{\bar{\sigma}} (\sigma_x + p); \quad d\gamma_{xy} = 3 \frac{d\bar{\epsilon}}{\bar{\sigma}} \tau_{xy} \\ d\epsilon_y &= {}^3/_2 \frac{d\bar{\epsilon}}{\bar{\sigma}} (\sigma_y + p); \quad d\gamma_{yz} = 3 \frac{d\bar{\epsilon}}{\bar{\sigma}} \tau_{yz} \\ d\epsilon_z &= {}^3/_2 \frac{d\bar{\epsilon}}{\bar{\sigma}} (\sigma_z + p); \quad d\gamma_{zx} = 3 \frac{d\bar{\epsilon}}{\bar{\sigma}} \tau_{zx} \end{aligned}\right\} \tag{4}$$

where the symbols have the following meaning:

$d\epsilon_x, d\epsilon_y, d\epsilon_z$ = incremental normal strain components in the x, y, z-co-ordinate directions, associated with the normal stress components

$\sigma_x, \sigma_y, \sigma_z$ = normal stress components, psi

$d\gamma_{xy}, d\gamma_{yz}, d\gamma_{zx}$ = incremental shear-strain components in plane normal to first subscript and in direction of second

$\tau_{xy}, \tau_{yz}, \tau_{zx}$ = shear-stress components, psi

$p = -{}^1/_3(\sigma_x + \sigma_y + \sigma_z)$ = mean hydrostatic pressure, psi

$$\bar{\sigma} = \left[\frac{(\sigma_x - \sigma_y)^2 + (\sigma_y - \sigma_z)^2 + (\sigma_z - \sigma_x)^2}{2} + 3(\tau_{xy}^2 + \tau_{yz}^2 + \tau_{zx}^2)\right]^{1/2}$$

= effective stress; it is proportional to the invariant of the deviatorial stress tensor; it is also the root-mean-square of shearing stresses, psi

$$d\bar{\epsilon} = {}^2/_3 \left[\frac{(d\epsilon_x - d\epsilon_y)^2 + (d\epsilon_y - d\epsilon_z)^2 + (d\epsilon_z - d\epsilon_x)^2}{2} + {}^3/_4(d\gamma_{xy}^2 + d\gamma_{yz}^2 + d\gamma_{zx}^2)\right]^{1/2}$$

= effective incremental strain

The quantity $d\bar{\epsilon}/\bar{\sigma}$ is also known as the plasticity modulus and relates stresses to strains not unlike Young's modulus in Hooke's law, except for the fact that the strains are strain increments and not small strains. The plasticity modulus $d\bar{\epsilon}/\bar{\sigma}$ is not a constant and depends on the state of work-hardening. Furthermore, total or finite strains can be constructed from the incremental strains of Equations (4) by the process of integration only as long as the strain or stress path is known. In the case of a simple stress path, as for instance in a simple tension test, when all stresses are equal to zero except one, Equations (4) can readily be evaluated. Let x be the axial direction in a tension test, then σ_x is the tension stress itself. Hence $\bar{\sigma}$ reduces to σ_x and the deviatorial stress $\sigma_x + p$ reduces to ${}^2/_3\sigma_x$. Substituting these into the first one of Equations (4) yields,

$$d\epsilon_x = d\bar{\epsilon}, \text{ or } \epsilon_x = \bar{\epsilon}$$

and similarly for second and third

$$\left.\begin{aligned} d\epsilon_y &= -{}^1/_2 d\bar{\epsilon}, \text{ or } \epsilon_y = -{}^1/_2 \bar{\epsilon} \\ d\epsilon_z &= -{}^1/_2 d\bar{\epsilon}, \text{ or } \epsilon_z = -{}^1/_2 \bar{\epsilon} \end{aligned}\right\} \tag{5}$$

while the remaining are zero.

Thus in a tension test, the effective stress-effective strain curve is the same as the true stress-true strain or natural-strain curve in tension. A diagrammatic $\bar{\sigma}$ versus $\bar{\epsilon}$ curve is shown in Fig. 5. The effective strain is assumed to be only the plastic strain, thereby ignoring the elastic contribution as being negligible except when spring-back determinations become important. Hence, in the region of uniform stressing below the maximum load (up to tensile strength) the axial finite strain ϵ_x or $\bar{\epsilon}$ is calculated by the natural logarithm of the ratio of final gage length l to initial gage length l_0; i.e., $\bar{\epsilon} = \ln (l/l_0)$. After necking, uniform deformation ceases to exist and $\bar{\epsilon}$ can no longer be calculated by $\ln (l/l_0)$. However, the ratio of initial and final diameters can now be used because of the constant volume relationship $l/l_0 = (D_0/D)^2$. Thus, $\bar{\epsilon} = 2 \ln (D_0/D)$ after necking. When yielding first begins, $\bar{\epsilon} = 0$ and $\bar{\sigma} = \sigma_0$ is equal to the yield strength in tension. As loading continues the $\bar{\sigma}$ versus $\bar{\epsilon}$ curve is traced as given in Fig. 5. Each point on this curve is then associated with a definite value of $\bar{\sigma}$ and $\bar{\epsilon}$.

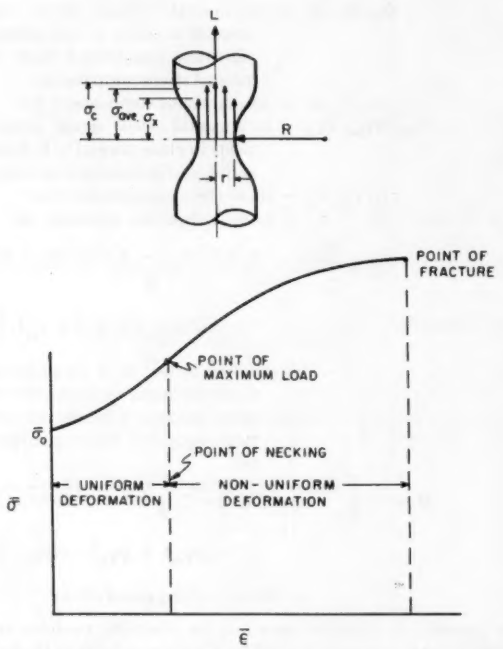


Fig. 5 Schematic sketch of a typical tensile effective stress-effective strain curve

It is important to realize that, at the instant when the maximum load is exceeded, the specimen begins to neck and a combined state of stress is initiated in the minimum section of the bar. As a matter of fact, continued straining will now only affect the necked region since the remainder of the test bar will no longer deform plastically. In order to calculate a true stress $\bar{\sigma}$ beyond the maximum load, it will not be possible to obtain it simply by dividing the load by the instantaneous minimum area in the neck. Bridgman [4] has shown, however, that a simple correction may be made in order to calculate the stress at the surface, providing the radius R of the neck contour is known. Thus, if L is the axial load, then

$$\sigma_x = \bar{\sigma} = \frac{L}{\pi r^2}\left(\frac{1}{1 + 2\frac{R}{r}\ln\left(1 + \frac{r}{2R}\right)}\right), \qquad (6)$$

in which it is assumed that the stress gradient will not affect the flow condition and that continued yielding will only take place when the surface stress $\bar{\sigma}$ is at the yield condition.

The amount of plastic work per unit volume at any particular strain $\bar{\epsilon}$ is given by the area under the curve, or

$$w = \int_0^{\bar{\epsilon}} \bar{\sigma} d\bar{\epsilon}, \qquad (7)$$

which is also called the specific energy. It is easy to see that the $\bar{\sigma}$ versus $\bar{\epsilon}$ curve is the trace of loci of instantaneous yield conditions and that at each locus $\bar{\sigma}$, $\bar{\epsilon}$, and w have a definite value. The important assumption in the theory is that, no matter in what manner a given state point is reached, the amount of specific energy required is always the same. Hence, the state of work-hardening is the same. This, of course, presupposes that the material is isotropic at all times. It is, however, well known that actual metals are not perfect in this respect and that they do actually exhibit various degrees of anisotropy depending on the deformation path and extent. It was found, for example, by Thomsen, et al. [10] that the correspondence among data obtained with compression, corrected tension, and torsion was fair for ductile aluminum alloys but poor for harder aluminum alloys. Hence any comparison that is to be made with metal-cutting data will be subject to these nonideal variations.

In order to relate the metal-cutting data to the $\bar{\sigma}$ versus $\bar{\epsilon}$ curve, it is necessary to assume a reasonable state of stress in the shear zone. The most likely one is that a maximum shearing stress exists on the shear plane and that the normal stresses are of equal magnitude; i.e., $-\sigma_x = -\sigma_y = -\sigma_z = p$. Hence a state of hydrostatic pressure exists on the shear plane and the deviator stresses $\sigma_x + p = \sigma_y + p = \sigma_z + p = 0$. Consequently, the three increments of the normal strain components of Equations (4) reduce to zero. Inasmuch as the z-direction may be assumed to coincide with the width direction and is a principal direction, therefore, $\tau_{yz} = \tau_{zx} = 0$ and only the shearing stress τ_{xy} remains as a nonzero shear stress. Hence, Equations (4) reduce to

$$d\gamma_{xy} = 3\frac{d\bar{\epsilon}}{\bar{\sigma}}\tau_{xy}$$

If the effective stress is now evaluated under the foregoing conditions, then $\bar{\sigma} = \sqrt{3}\,\tau_{xy}$ and substituting and integrating results in

$$\bar{\epsilon} = \frac{\gamma_{xy}}{\sqrt{3}}, \quad \text{when} \quad \bar{\sigma} = \sqrt{3}\,\tau_{xy} \qquad (8)$$

It is clear now that ideally the specific energy in a tension test (or compression test) would yield the same state of work-hardening in a static metal-cutting test (carried out infinitely slowly), whenever

$$w = \int_0^{\bar{\epsilon}} \bar{\sigma} d\bar{\epsilon} = \int_0^{\gamma_{xy}} \sqrt{3}\tau_{xy}\frac{d\gamma_{xy}}{\sqrt{3}} = \int_0^{\gamma_{xy}} \tau_{xy} d\gamma_{xy} \qquad (9)$$

where the strains and stresses are related by Equations (4). In view of the fact that anisotropy may develop during deformation and because of the fact that the state of stress is simplified and does not hold at the intersection of the shear plane with the surfaces of the workpiece being machined, perfect agreement cannot be expected.

The foregoing theory was first applied to metal-cutting data by Lapsley, Grassi, and Thomsen [11]. These authors found that their correlation with tensile data was moderately successful, in spite of the fact that an intercept correction, as discussed earlier, was not applied. Marshall and Shaw [12] recognizing defi-

ciency in the correlation proposed that, on the basis of representing the state of stress and of small strain by Mohr's circle, a better correlation could be achieved. This artifice, however, is not necessary, for the relationship used by Marshall and Shaw can be obtained directly from the plasticity Equations (4) by merely assuming that the effective stress under certain circumstances may have a different value. It is well known that for certain materials the maximum shear stress is a better criterion than the effective stress obtained from the distortion-energy theory. It is generally believed, for example, that the distortion-energy theory fits better for ductile metals, while the maximum-shear-stress theory tends to give better agreement with brittle materials. Thus substituting $\bar{\sigma} = 2\tau_{xy}$ in the fourth equation and integrating yields $\gamma_{xy} = \frac{3}{2}\bar{\epsilon}$. Hence the two criteria become:

$$\left.\begin{array}{l} 1 \quad \text{Distortion-energy theory} \\ \qquad \bar{\sigma} = \sqrt{3}\,\tau_{xy}, \quad \bar{\epsilon} = \dfrac{\gamma_{xy}}{\sqrt{3}} \\ 2 \quad \text{Maximum-shear-stress theory} \\ \qquad \bar{\sigma} = 2\tau_{xy}, \quad \bar{\epsilon} = \frac{3}{2}\gamma_{xy} \end{array}\right\} \tag{10}$$

It is to be noted that hypothesis 2 for the same shearing stress and strain would yield an increased specific energy over that of hypothesis 1.

Fig. 6 shows, superimposed on the compression stress-strain curves for steel SAE 4135, the data points of tension tests to which the Bridgman correction, after the initiation of necking, was applied. During the tests the instantaneous neck-contour diameters were obtained together with the neck diameter by fitting a cone to the neck surface in a manner similar to that reported by Marshall and Shaw [12]. It is at once evident that agreement is good but not perfect. However, inasmuch as the deviation occurs at small strains, the choice of one test over the other would result in little or no error. Fig. 7 shows comparison of tension with compression data for mild steel and an aircraft aluminum alloy.

Before discussing the experimental metal-cutting data points in the light of the plastic-flow theory, we may examine the several possibilities for correlation. On account of what has been said earlier about the possible mechanism at high rates of strain in metal cutting, it is not likely that the stress rises during straining in accordance with the assumed stress-strain curve $\overline{fg}$ of Fig. 8(*a*). In fact it is much more likely that the path $\overline{ag}$ would represent the actual sequence of events and hence the specific energy would be $A_1 + A_2 = \tau_{xy}\gamma_{xy}$ rather than

$$A_1 = \int_0^{\gamma_{xy}} \tau_{xy}\,d\gamma_{xy}$$

Thus, if the point g represents a metal-cutting data point which correlates with the property test, then the work of deformation between metal cutting and the static tension test are not the same. Bastien and Weisz [13] suggest that correlation is achieved by comparing specific energies, which would demand in the present case that point g fall above the static-test curve for the same strain. It also may be noted that if the correlation were based on the maximum-shear-stress theory, the metal-cutting data point would be at h. If correlation is to be achieved, therefore, it will be necessary that the stress-strain curve be that given by $\overline{ih}$. In the sketch of Fig. 8(*b*) is given a comparison of the dynamic energy based on areas equivalent to $A_1 + A_2$ for every $\bar{\sigma}$, while the static curve is the area under the stress-strain curve. A similar comparison is made in Fig. 8(*c*) in terms of stresses, and it is to be noted that there is essentially no difference in the representation of these two energies in Figs. 8(*b*) and 8(*c*).

The experimental data points for all test conditions investigated, Tables 1 to 5, are plotted on the $\bar{\sigma}$ and $\bar{\epsilon}$ logarithmic coordinates based on hypothesis 1; i.e., the distortion-energy principle

$$\bar{\sigma} = \sqrt{3}\,\tau_{xy} \quad \text{and} \quad \bar{\epsilon} = \frac{\gamma_{xy}}{\sqrt{3}}$$

in Figs. 9 to 11, inclusive. All experimental data were corrected by subtracting the intercept values from the F_s versus A_s curves for the particular alloy in question. It is seen that the correlation with $\bar{\sigma}$ for approximately the same strain $\bar{\epsilon}$ is good for the free-cutting steel SAE 1112, annealed, as-received, and steel SAE 1020 as-received, Fig. 9, and for aluminum alloy 6061-0, 6061-T6, and alpha-brass, Fig. 11. It is to be noted that the materials which were tested previously at intermediate speeds

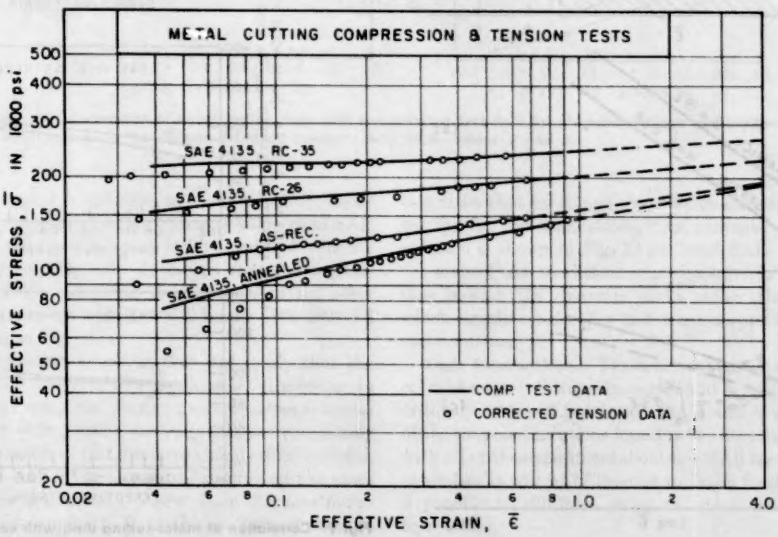


Fig. 6 Compression and corrected tension effective stress-effective strain data for steel SAE 4135. Bridgman correction applied to tension test data.

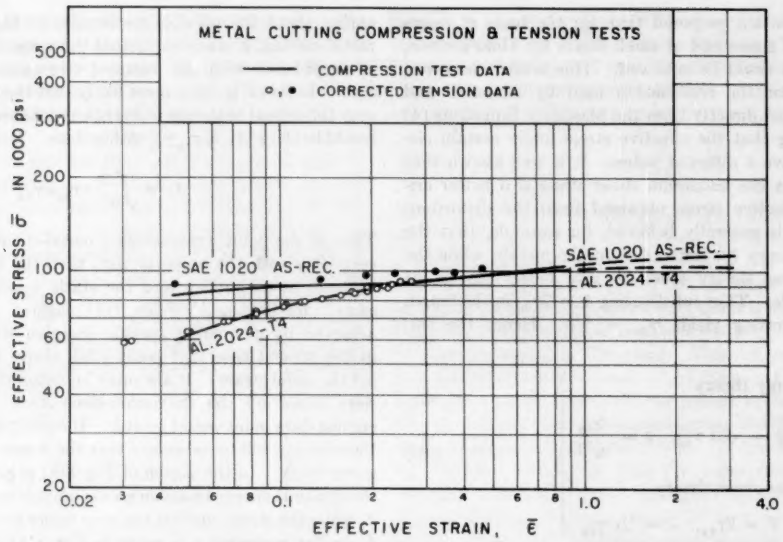


Fig. 7 Compression and corrected tension effective stress-effective strain data for steel SAE 1020 as-received and aluminum alloy 2024-T4

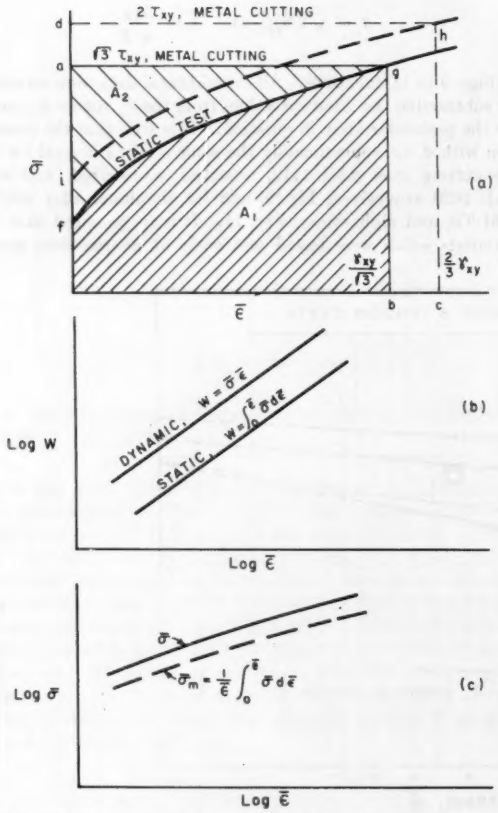


Fig. 8 Schematic representation of dynamic and static stresses and specific energy

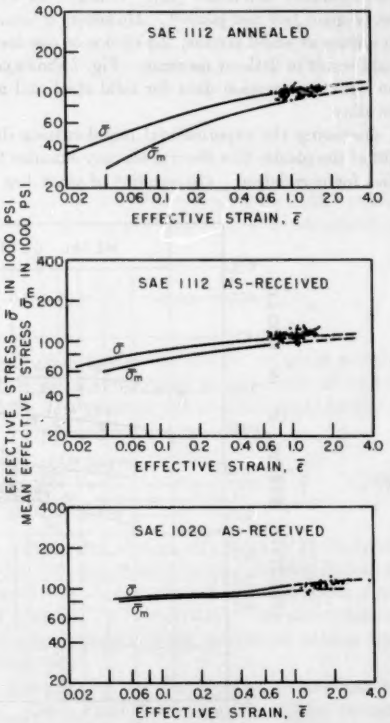

SAE 1112 ANNEALED
SAE 1112 AS-RECEIVED
SAE 1020 AS-RECEIVED
EFFECTIVE STRESS $\bar{\sigma}$ IN 1000 PSI
MEAN EFFECTIVE STRESS $\bar{\sigma}_m$ IN 1000 PSI
EFFECTIVE STRAIN, $\bar{\epsilon}$

Fig. 9 Correlation of metal-cutting data with compression test data for steel SAE 1112 and steel SAE 1020, based on the distortion-energy theory. Data from Tables 1 and 2.

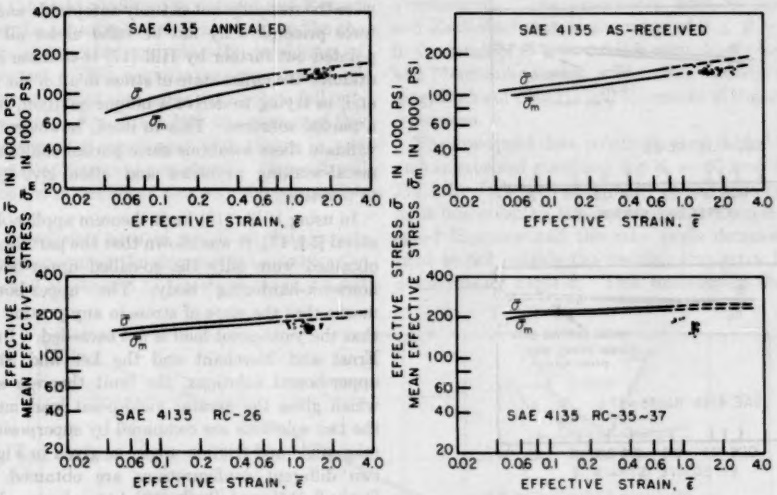

SAE 4135 ANNEALED
SAE 4135 AS-RECEIVED
SAE 4135 RC-26
SAE 4135 RC-35~37
EFFECTIVE STRESS $\bar{\sigma}$ IN 1000 PSI
MEAN EFFECTIVE STRESS $\bar{\sigma}_m$ IN 1000 PSI
EFFECTIVE STRAIN, $\bar{\epsilon}$

Fig. 10 Correlation of metal-cutting data with compression test data for alloy steel SAE 4135, based on the distortion-energy theory. Data from Table 3.

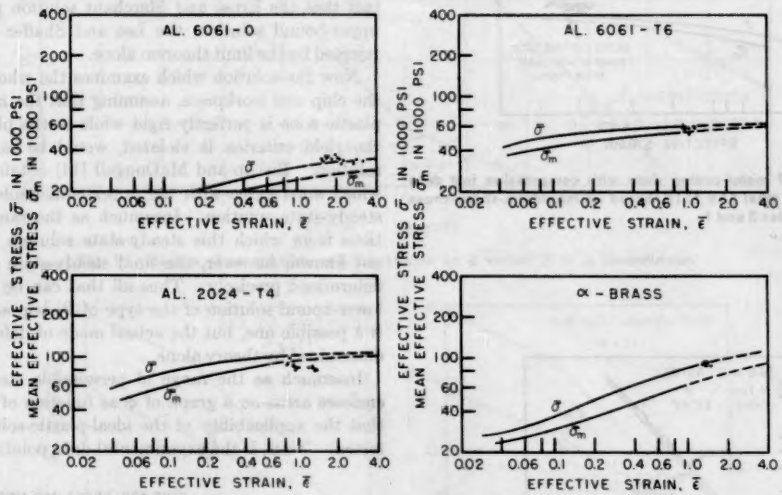


Fig. 11 Correlation of metal-cutting data with compression test data for aluminum alloys and alpha-brass based on the distortion-energy theory. Data from Tables 4 and 5.

[1, 3] show the same good correlation as those presently under discussion. All four metallurgical structures of alloy steel SAE 4135, however, yield data points which fall below the $\bar{\sigma}$ versus $\bar{\epsilon}$ curve in varying degrees, Fig. 10. It is seen that the correlation becomes increasingly poorer as the hardness of the alloy increases. Similar imperfect correlation is found with 2024-T4 aluminum alloy shown in Fig. 11.

The imperfect correlation, however, does not imply that the maximum-shear-stress theory is now applicable. Referring to Fig. 12, it is seen that using the maximum-shear-stress criterion for plastic flow, i.e., $\bar{\sigma} = 2\tau_{xy}$ and $\bar{\epsilon} = {}^2/_3\gamma_{xy}$, does not completely cure all evils. As a matter of fact, for a reasonably ductile metal, as for example steel SAE 1112 as-received, shown in the same figure, the correlation is definitely poorer when the maximum-shear-stress criterion is applied. It is also of interest to note that a graphical representation, in terms of energies suggested by Bastien and Weisz [13], Fig. 8(*b*), gives no different correlation than that by comparing $\bar{\sigma}$ in tension with $\sqrt{3}\,\tau_{xy}$ in metal cutting for the same strain. An example of this type of presentation is shown in Fig. 13 for steel SAE 1112 as-received for the several test conditions investigated. It is seen that correlation is with the dynamic curve rather than the static curve which simply means that $\bar{\sigma}$ in tension correlates with $\sqrt{3}\,\tau_{xy}$ in metal cutting.

Angle Relationships. The functional angle relationship between α (rake angle), β (friction angle) and ϕ (shear-plane angle) were critically reviewed previously [1, 2] and it was found that none of the proposed solutions found in the literature was in agreement with all of the experimental data. With the additional data now available, it will be of interest to make further comparisons and if possible to eliminate solutions which fail to yield reasonable correlation.

It was pointed out by Bishop and McDougall [14] that the Ernst and Merchant [15] and the Lee and Shaffer [16] solutions

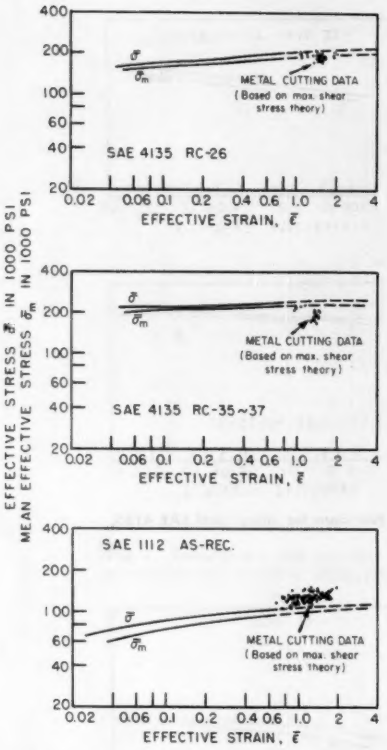

EFFECTIVE STRESS $\bar{\sigma}$ IN 1000 PSI
MEAN EFFECTIVE STRESS $\bar{\sigma}_m$ IN 1000 PSI
SAE 4135 RC-26
METAL CUTTING DATA
(Based on max. shear stress theory)
EFFECTIVE STRAIN, $\bar{\epsilon}$
SAE 4135 RC-35~37
SAE 1112 AS-REC.

Fig. 12 Correlation of metal-cutting data with compression test data for steel SAE 4135 and steel SAE 1112, based on maximum-shear-stress theory. Data from Tables 3 and 1.

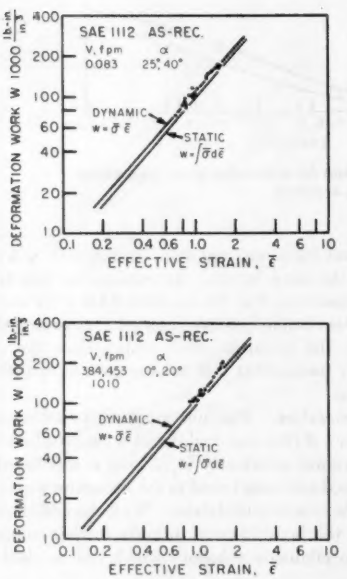


Fig. 13 Plastic deformation work as a function of effective strain for steel SAE 1112 as-received, based on distortion-energy theory

were theoretically not entirely acceptable and that the minimum-force principle may not be valid under all conditions. It was pointed out further by Hill [17] that these investigators did not examine the entire state of stress in all of the metal-cutting region and, in trying to derive a unique solution, indeed obtained only a partial solution. This in itself, however, is insufficient to invalidate these solutions since partial solutions are used in other metal-working problems and often give good approximations to practice.

In using Hill's [18] limit theorem applicable to an ideal plastic metal [14, 17], it was shown that the partial solutions which were obtained were only the so-called upper-bound solutions for a nonwork-hardening body. The upper-bound solution here means that the state of stress in any portion of the metal is such that the yield-point load is not exceeded. Inasmuch as both the Ernst and Merchant and the Lee and Shaffer solutions are upper-bound solutions, the limit theorem states that the one which gives the greater yield-point load must be rejected. If the two solutions are compared by superposition for a particular rake angle and friction angle, as given in Fig. 14, it is seen that two different configurations are obtained. Bishop and McDougall [14] and Pugh [19] have discussed this situation and state that neither one of the solutions can be rejected on the basis of yield-point load, since neither one has been found to be applicable from the inception of cutting. Thus, in spite of the fact that the Ernst and Merchant solution would give a better upper-bound solution, the Lee and Shaffer solution cannot be rejected by the limit theorem alone.

Now the solution which examines the whole state of stress in the chip and workpiece, assuming that the metal outside of the plastic zone is perfectly rigid while in the plastic zone nowhere the yield criterion is violated, would be called a lower-bound solution. Bishop and McDougall [14] obtained such a solution, which must agree with the upper-bound solution for a possible steady-state solution. Inasmuch as the range of initial conditions from which this steady-state solution was constructed is not known, however, the final steady-state solution cannot be determined precisely. Thus all that can be said now is that a lower-bound solution of the type of Bishop and McDougall [14] is a possible one, but the actual mode of deformation cannot be determined by theory alone.

Inasmuch as the range of permissible steady-state solutions encloses areas on a graph of ϕ as function of $\beta - \alpha$, it appears that the applicability of the ideal-plastic-solid solutions can be tested. Thus, if the experimental data points fall outside of this

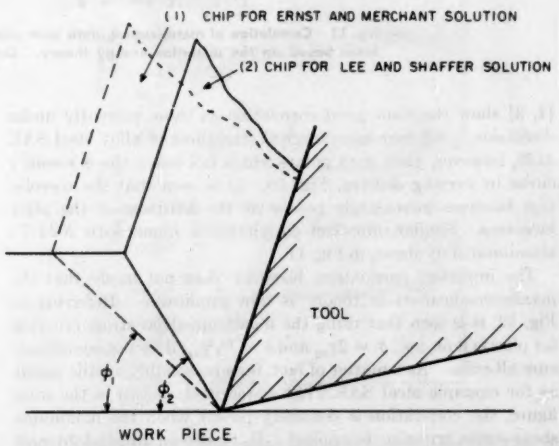


Fig. 14 Two possible upper-bound solutions for the shear plane at identical α and β

permissible area, then the yield condition is violated and it must be assumed that either the actual metal does not behave like the ideal metal or the model used for analyzing the mechanics of metal cutting is incorrect. This examination was made recently [1, 2, 20] and it was found that some data points fell outside of the permissible area. Also Pugh [19] has examined this and finds that the agreement of experiment with theory is poor when all of the limiting conditions are considered. Hence it will be of benefit to examine the present data also with the permissible area.

Fig. 15 shows the permissible area in which all data points must lie in order that the metal at the singularities given by the apex of the angles δ, γ, θ at the end of the shear plane are not overstressed. This permissible area is bounded by the Lee and Shaffer solution in the range of $0 \leq \beta - \alpha \leq 45$ deg, while in the range of $\beta - \alpha < 0$ it is very closely bounded by the Ernst and Merchant solution. The other boundaries are given by the magnitudes of β and α and the extent of the area reduces as β and α increase.

The low-speed data points for steel SAE 1112 in the annealed and as-received condition for $\alpha = 40$ and 25 deg are given in Fig. 16. It is seen that many of the data points fall below the limit line given by the overstressing of angle γ. As the cutting speed increases and the rake angle decreases, the data points tend to fall outside the limiting line given by the condition of overstressing angle θ. This tendency is shown in Fig. 17 for

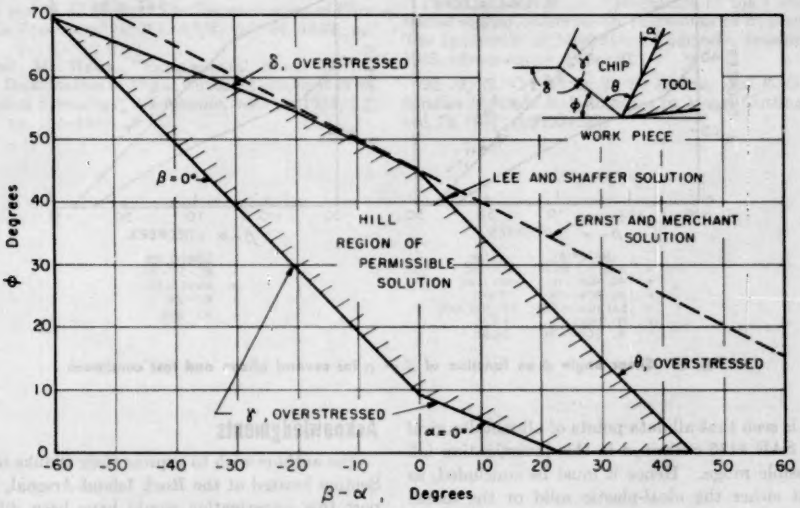


Fig. 15 Region of permissible solution shown on ϕ versus $\beta - \alpha$ co-ordinates

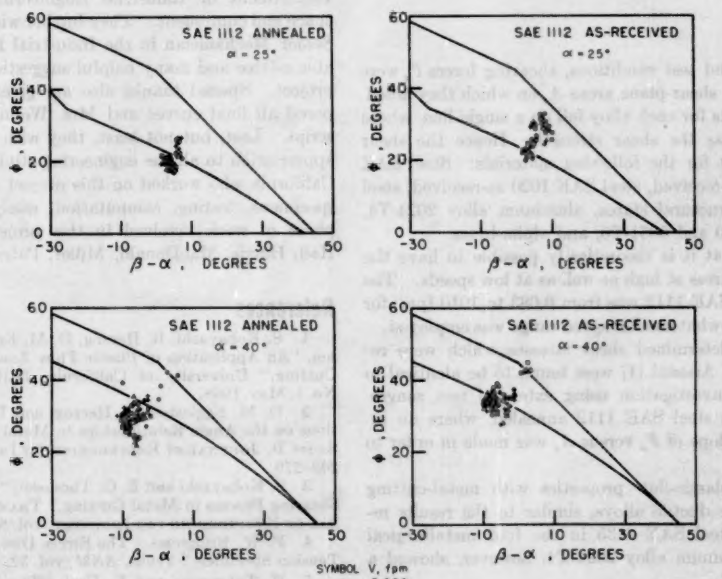


Fig. 16 Shear angle ϕ as function of $\beta - \alpha$ under various test conditions for steel SAE 1112 annealed and as-received

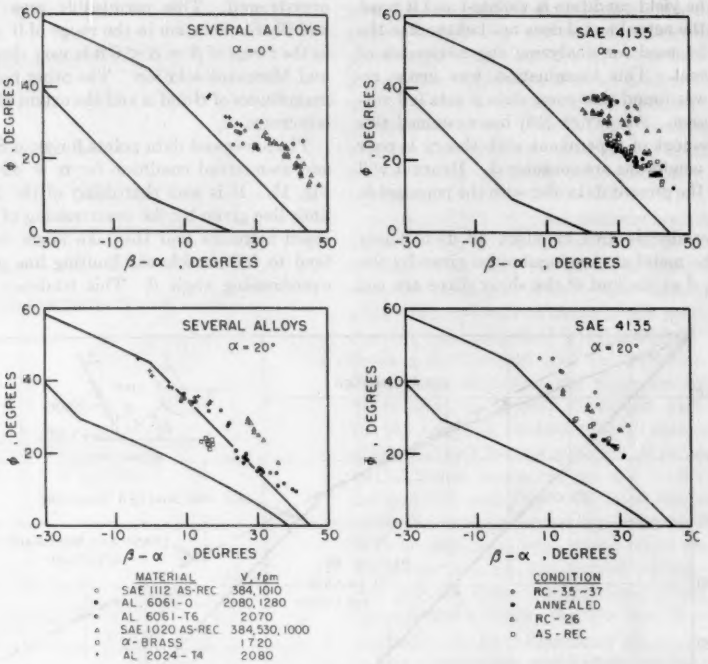


Fig. 17 Shear angle ϕ as function of $\beta - \alpha$ for several alloys and test conditions

several alloys. It is seen that all data points of all metallurgical structures of steel SAE 4135 examined in this investigation fall beyond the permissible range. Hence it must be concluded, as stated before, that either the ideal-plastic solid or the model representing the metal-cutting process (or both) are perhaps not valid assumptions.

Conclusions

1 For all alloys and test conditions, shearing forces F_s were linear functions of the shear-plane areas A_s on which they acted. All experimental points for each alloy fell on a single line, whose slope was expressed as the shear strength. Hence the shear stresses were constant for the following materials: Steel SAE 1112 annealed and as-received, steel SAE 1020 as-received, steel SAE 4135 in four structural states, aluminum alloy 2024-T4, aluminum alloy 6061-0 and 6061-T6, and alpha-brass.

2 It was shown that it is theoretically possible to have the same constant shear stress at high as well as at low speeds. The speed range for steel SAE 1112 was from 0.083 to 1010 fpm; for the other alloys a somewhat smaller speed range was employed.

3 All previously determined shear stresses which were reported to Rock Island Arsenal [1] were found to be identical to those found in this investigation using extended test ranges. The only exception is steel SAE 1112 annealed, where an insignificant change of slope of F_s versus A_s was made in order to improve correlation.

4 Correlation of plastic-flow properties with metal-cutting data was good for the ductile alloys, similar to the results reported earlier [1]. Steel SAE 4135 in the four metallurgical states as well as aluminum alloy 2024-T4, however, showed a less perfect correlation.

5 The three important theoretical angle relationships of Ernst and Merchant, Lee and Shaffer, and Hill, were found to be in incomplete agreement with the data presented in this paper.

Acknowledgments

The authors wish to express their thanks to the Machinability Section located at the Rock Island Arsenal, without whose support this investigation would have been difficult to carry out. They also wish to thank the University of California and the Department of Industrial Engineering for making available space and equipment. They further wish to thank Mr. Petersen, Senior Mechanician in the Industrial Engineering shops, for his able advice and many helpful suggestions during the life of the project. Special thanks also are given to Mr. Kot, who prepared all final curves and Mrs. Wehner for typing the manuscript. Last, but not least, they wish to express their sincerest appreciation to all the engineering students at the University of California who worked on this project in the form of preparing specimens, testing, computation, analyzing—in fact, in every phase of work involved in this project. These students are Hall, Harris, MacDonald, Miller, Pabor, Saylor, and Thometz.

References

1 S. Kobayashi, R. Herzog, D. M. Eggleston, and E. G. Thomsen, "An Application of Plastic Flow Analysis to Orthogonal Metal Cutting," University of California, IER Report No. 117, Issue No. 1, May, 1958.

2 D. M. Eggleston, R. Herzog, and E. G. Thomsen, "Observations on the Angle Relationships in Metal Cutting," TRANS. ASME, Series B, JOURNAL of ENGINEERING for INDUSTRY, vol. 81, 1959, pp. 263–279.

3 S. Kobayashi and E. G. Thomsen, "Some Observations on the Shearing Process in Metal Cutting," TRANS. ASME, Series B, JOURNAL OF ENGINEERING FOR INDUSTRY, vol. 81, 1959, pp. 251–262.

4 P. W. Bridgman, "The Stress Distribution at the Neck of a Tension Specimen," *Trans. ASM*, vol. 32, 1944, pp. 553–574.

5 H. Takeyama and E. Usui, "The Effect of Tool-Chip Contact Area in Metal Cutting," TRANS. ASME, vol. 80, 1958, pp. 1089–1096.

6 G. Vieregge, "Temperaturfeld und Wärmebilanz des Schervorganges bei der Zerspanung," *Werkstatt und Betrieb*, vol. 88, 1955, pp. 227–230.

7 A. H. Cottrell, "Deformation of Solids at High Rates of Strain," Proceedings of the Conference on the Properties of Materials at High Rates of Strain, The Institution of Mechanical Engineers, London, England, April-May, 1957, pp. 1–22.

8 Z. S. Basinski, "The Instability of Plastic Flow of Metals at Very Low Temperatures," *Proceedings of the Royal Society*, series A, vol. 240, 1957, pp. 229–242.

9 B. T. Chao and K. J. Trigger, "Controlled Contact Cutting Tools," TRANS. ASME, Series B, JOURNAL OF ENGINEERING FOR INDUSTRY, vol. 81, 1959, pp. 139–151.

10 E. G. Thomsen, I. Cornet, I. Lotze, and J. E. Dorn, "Investigation on the Validity of an Ideal Theory of Elasto-Plasticity for Wrought Aluminum Alloys," NACA TN 1552, 1948.

11 J. T. Lapsley, Jr., R. C. Grassi, and E. G. Thomsen, "Correlation of Plastic Deformation During Metal Cutting With Tensile Properties of the Work Material," TRANS. ASME, vol. 72, 1950, pp. 979–986.

12 E. R. Marshall and M. C. Shaw, "The Determination of Flow Stress From a Tensile Specimen," *Trans. ASM*, vol. 44, 1952, pp. 705–720.

13 P. Bastien and M. Weisz, "Experimental Study of the Mechanism of Plastic Deformation in Metal Cutting Compared With the Methods of Classical Stressing," *Microtecnic*, vol. 11, 1957, pp. 122–129; vol. 11, 1957, pp. 151–161.

14 J. F. W. Bishop and W. S. McDougall, "A Note on Machining Under Plane Strain Conditions," Department of Scientific and Industrial Research, Mechanical Engineering Laboratory, East Kilbride, Glasgow, Scotland, MF/MS 74, 1953.

15 H. Ernst and M. E. Merchant, "Chip Formation, Friction, and High Quality Machined Surfaces," from "Surface Treatment of Metals," *Trans. ASM*, vol. 29, 1941, pp. 299–328.

16 E. H. Lee and B. W. Shaffer, "The Theory of Plasticity Applied to a Problem of Machining," *Journal of Applied Mechanics*, vol. 18, TRANS. ASME, vol. 73, 1951, pp. 405–413.

17 R. Hill, "The Mechanics of Machining: A New Approach," *Journal of the Mechanics and Physics of Solids*, vol. 3, 1954, pp. 47–53.

18 R. Hill, "On the Limits Set by Plastic Yielding to the Intensity of Singularities of Stress," *Journal of the Mechanics and Physics of Solids*, vol. 2, 1954, pp. 278–285.

19 H. Ll. D. Pugh, "Mechanics of the Cutting Process," Presented at the Conference on Technology of Engineering Manufacture, The Institution of Mechanical Engineers, London, England, March, 1958, advance copy, Paper 53.

20 J. H. Creveling, T. F. Jordan, and E. G. Thomsen, "Some Studies of Angle Relationships in Metal Cutting," TRANS. ASME, vol. 79, 1957, pp. 127–138.

New Developments in the Theory of the Metal-Cutting Process

Part I. The Ploughing Process in Metal Cutting

P. ALBRECHT
Research Associate,
Physical Research Department,
The Cincinnati Milling Machine Company,
Cincinnati, Ohio

Revelation of the significance of "ploughing" in the metal-cutting process, which occurs because of the finite sharpness of the cutting edge, leads to a better understanding of the mechanics of the metal-cutting process. The concept of the ploughing force on the extreme cutting edge allows the development of a more complete force diagram which separates the ploughing force from the chip-tool interface force. Components of this more detailed force diagram have been verified experimentally. In terms of the new force diagram the real value of the coefficient of friction on the chip-tool interface has been found and the paradox of variation of the coefficient of friction with variation of rake angle explained.

The paper also contributes to a better understanding of such events as the effect of cutting velocity upon tool forces, built-up edge, chip curling, and residual stresses in the work surfaces.

Introduction

Theory of Metal-Cutting Process. The theory of the metal-cutting process has been considerably developed since the end of the last century. The principal mechanism in chip formation has been recognized to be a shearing process taking place along a shear plane in the work material. By the shearing action the work material is plastically deformed and separated from the workpiece. Usually the thickness of the chip is more than the "thickness of bite." In order to produce such plastic deformation of the metal considerable forces are required. These forces acting between the cutting tool and work material are shown in the force diagram, Fig. 1. Although the theory, as it is presently known, clarifies to a great extent the mechanics of the metal-cutting process, unexplained dark corners remain.

To name a few such instances, we may note here that the coefficient of friction on the chip-tool interface has usually been found to be rather high compared with values found by direct experiment on surfaces having contaminating films. Further, this coefficient of friction, as given by the theory, has a peculiar property, i.e., to vary with changing tool geometry. The chip-tool coefficient of friction increases, namely, with increasing rake angle of the cutter. These and other unexplained results have expressed the need for research and further development of the theory of the metal-cutting process.

Because it was felt by the author that the theory of the metal-cutting process based on the analytical description of the shearing process alone had reached its maximum development, it was hoped that further progress could be achieved by introducing into the theory some other mechanisms which may be present in the metal-cutting process and which, together with shearing, affect its behavior. Generally, the metal-cutting process, as well as any

Contributed by the Production Engineering Division and presented at the Annual Meeting, Atlantic City, N. J., November 29–December 4, 1959, of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS.

NOTE: Statements and opinions advanced in papers are to be understood as individual expressions of their authors and not those of the Society. Manuscript received at ASME Headquarters, August 14, 1959. Paper No. 59—A-243.

Nomenclature

Variables of tool shape (for the orthogonal cutting process)

- r = sharpness radius, which gives magnitude of sharpness rounding of cutting edge
- r_e = effective sharpness radius, which accounts for change of sharpness rounding due to presence of the built-up edge
- α = rake angle
- δ = included angle of cutting edge

Dimensions of Cut

- t_1 = thickness of uncut chip
- t_2 = thickness of chip after its separation from the work
- w = width of cut in orthogonal cutting

Ratios of Cutting Process

- $r_t = t_1/t_2$ = chip thickness ratio
- $\lambda = r/t_1$ or r_e/t_1 = ploughing ratio, for cutting process without or with built-up edge

Tool Forces

- R = resultant tool force (or briefly tool force)
- F_c = cutting force (component of R in direction of tool motion)
- F_t = thrust force (component of R perpendicular to direction of tool motion)
- P = main ploughing force or chip ploughing force at extreme cutting edge
- P_1 = component of force P in direction of tool motion
- P_2 = component of force P perpendicular to direction of tool motion
- P_n = normal component of force P
- P_f = friction component of force P
- P_s = component of ploughing force polygon which takes part in shearing
- Q = tool face force

(*Continued on next page*)

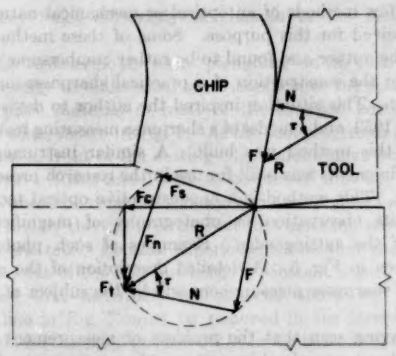


Fig. 1 Conventional force diagram of orthogonal cutting processes

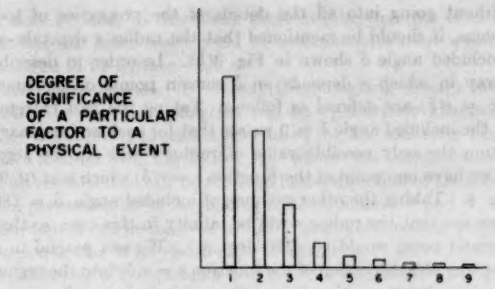

DEGREE OF SIGNIFICANCE OF A PARTICULAR FACTOR TO A PHYSICAL EVENT
1 2 3 4 5 6 7 8 9
NUMBER OF EACH FACTOR AS CLASSIFIED ACCORDING TO ITS SIGNIFICANCE IN A PHYSICAL EVENT

Fig. 2 Significance of various factors to a physical event

other physical event, can be expected to depend on a number of factors which will affect it to different degrees, as is symbolically shown in Fig. 2. The main factor in the metal-cutting process has been recognized as the shearing process in chip formation. Now, as a second mechanism taking place in metal cutting, the ploughing process is suggested here, the description of which and introduction into the theory of metal cutting is the purpose of this paper.

It will be shown in this paper that, by taking into account ploughing, metal-cutting theory is given more power to explain experimental results which are virtually impossible to interpret when the influence of this factor is omitted. For this reason ploughing is believed to be the number two mechanism in metal cutting and introduction of it into the theory makes the analytical model of the metal-cutting process more dependable in prediction and more powerful in explanation of experimental results.

Sharpness of Metal-Cutting Tools. The sharpness of metal-cutting tools should be discussed briefly at this point because it will be an essential ingredient of the theory developed in the main part of this paper. Often tool sharpness in metal cutting is associated with tool wear. Thus if the flank or crater wear of a cutting tool is in an advanced stage the tool is often described as blunt or dull. In this paper, the term sharpness will be understood to define the very small rounding of the extreme cutting edge, the tiny approximately cylindrical surface connecting tool-flank and tool-face surfaces.

This small cylindrical surface is developed on the cutting edge during the grinding process of the cutting tool. The way in which it happens is illustrated in Fig. 3(*a*). As the grinding wheel works along the tool face or tool flank, the tiny particles of tool material will break off at the extreme edge where the edge is so thin that

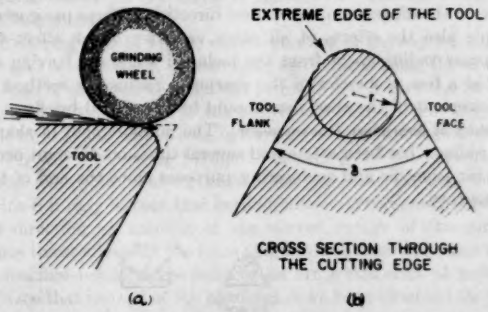


Fig. 3(*a*) In grinding of a cutter the wheel breaks off small particles at the extreme edge. This results in a rounded surface between the flank and face of a tool. (*b*) The radius *r* of the circle approximating the cross section of the rounded extreme edge is a measure of the sharpness of the cutting edge. The included angle δ of the cutting edge is also shown.

the material cannot stand the impact of grinding-wheel grains. This leads to a rounded surface as shown in Fig. 3. It should be noted here that the grinding wheel does not generate the rounded surface and its surface finish can be expected to be much rougher than those of the surfaces which were produced by the grinding wheel directly.

The cross section of this rounding would be a curve which can be very well approximated by a segment of a circle as shown in Fig. 3(*b*). The radius *r* of such a circle gives the magnitude of the rounding and thus can be adopted as a measure of the sharpness of the tool. The magnitude of the radius characterizing the tool sharpness depends on a number of factors such as tool material, method of grinding, and some others.

Nomenclature

Q_1 = component of force Q in direction of tool motion
Q_2 = component of force Q perpendicular to direction of tool motion
Q_n = normal component of force Q
Q_f = frictional component of force Q
Q_s = component of Q in direction of shear force S_s
P' = work ploughing force. This force can be resolved into components in a way similar to force P yielding the components P_1', P_2', P_n', and P_f'
L = land force; force on the wear land of the tool flank
L_1 = component of force L in direction of tool motion, or frictional component of force L
L_2 = component of force L perpendicular to direction of tool motion, or normal component of the force L
S = total force on shear plane
S_s = component of S in direction of shearing at shear plane

Shear Plane Position

ϕ = shear angle

Workpiece Material Properties

τ_s = mean shear stress on shear plane

Coefficients of Friction and Friction Angles

$\mu_Q = Q_f/Q_n = \tan \tau_Q$ = coefficient of friction on chip-tool interface
τ_Q = friction angle corresponding to coefficient of friction μ_Q
$\mu_r = P_f/P_n = \tan \tau_r$ = coefficient of friction on sharpness rounding of cutting edge
τ_r = friction angle corresponding to coefficient of friction μ_1
$\mu_L = L_1/L_2 = \tan \tau_L$ = coefficient of friction on wear land of tool flank
τ_L = friction angle corresponding to coefficient μ_L

Without going into all the details of the properties of tool sharpness, it should be mentioned that the radius r depends on the included angle δ shown in Fig. 3(*b*). In order to describe the way in which r depends on δ certain points of the function $r = r(\delta)$ are defined as follows: Taking the first extreme when the included angle $\delta = 0$ we see that for such an imaginary situation the only possible value of radius r will also be zero. Thus we have one point of the function $r = r(\delta)$ which is at (0, 0) in Fig. 4. Taking the other extreme of included angle, $\delta = 180$ deg, we see that the radius r will be infinity in this case, so that the second point would be (180 deg, ∞). We can extend in a similar way this discussion of the function $r = r(\delta)$ into the region of negative sharpness, that is for values of included angle greater than 180 deg. Results of the whole discussion are given in Fig. 4. The character of the graph indicates that the function $r = r(\delta)$ may be approximated by the analytical expression $r = a \tan^n(\delta/2)$ where a and n are parameters which determine the position of the curve on the diagram in the vertical direction. These parameters include also the effects of all other variables which affect the sharpness radius apart from the included angle δ. Having indicated a few properties of the sharpness radius the method of measurement of the sharpness should be mentioned briefly.

Method of Sharpness Measurement. The measurement of sharpness radius r has been attempted several times on different occasions for scientific and engineering purposes since the end of the last century.

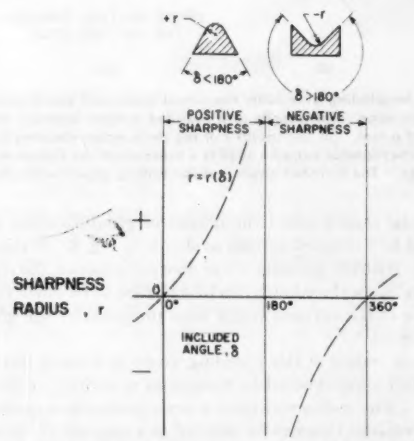


Fig. 4 General character of the relation between the sharpness radius r and included angle of the cutting edge

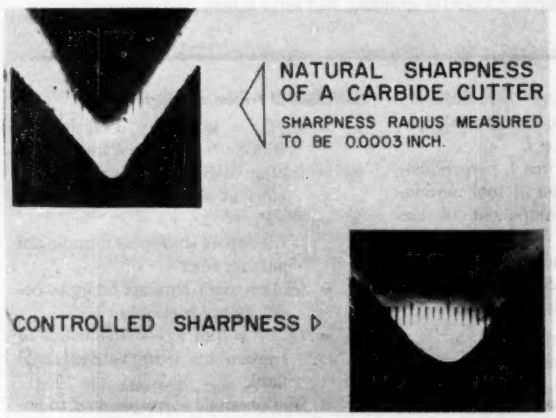


Fig. 5 Sharpness measurement by the method of optical cross sections

Only a few methods of an optical or mechanical nature have been conceived for this purpose. Some of these methods were tried by the author and found to be rather cumbersome and unsuitable for the construction of a practical sharpness-measuring instrument. This situation inspired the author to devise a new method in 1953, and a model of a sharpness measuring instrument based on this method was built. A similar instrument with further refinements was built for use in the research presented in this paper. This method is a nondestructive optical technique, and permits observation or photography of magnified cross sections of the cutting edge. Examples of such photographs may be seen in Fig. 5. A detailed description of the author's method of sharpness measurement will be the subject of a separate paper.

After having seen that the problem of measurement of tool sharpness has been solved in a practical way we proceed to the main part of this paper where use will be made of the notion of the tool sharpness as well as of sharpness measurement.

Theoretical Considerations

Ploughing in the Metal-Cutting Process. It has been the principal purpose of this work to take into account explicitly the process of ploughing which occurs in every metal-cutting process due to the finite sharpness of the cutting tool. This should be described now in greater detail.

The usual picture of the metal-cutting process is shown at the left in Fig. 6. Here the tool is considered as being infinitely sharp by disregarding the small rounding on the extreme cutting edge. Due to that simplification, the whole chip-tool pressure has a constant direction, provided the coefficient of friction does not vary along the tool face, and can be summed up in one resultant force of the same direction. This has been the usual assumption for the development of the theory of the metal-cutting process. If it is decided to take into account the fact that the tool has a finite sharpness, it is necessary to add to the foregoing concept a rounding on the extreme cutting edge as shown at the right in Fig. 6. Just as on the tool face, pressure will occur along the rounded portion, but the direction of this pressure is different from that on the tool face and varies from point to point. Also, the coefficient of friction on the rounded surface of the extreme edge may be expected to have a different value from that on the tool face, especially if a built-up edge covers this portion of the tool as is usually the case. Because of this rounding on the extreme cutting edge it will be much more enlightening if the entire system of chip-tool pressures, including those on the rounded cutting edge, are not summed up into one resultant force. Rather, this paper will sum up the pressures along the straight portion of the tool face into one resultant Q, and the pressures on the rounded portion of the extreme edge into another resultant P, which will be designated the ploughing force. Later it will be seen that these two forces should be kept separate in development of the theory be-

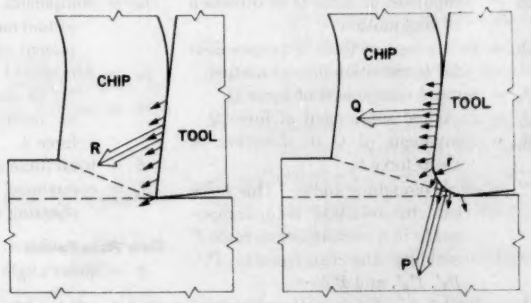


Fig. 6 The resolving of the resultant tool force R into two components; ploughing force P and tool face force Q

cause they react differently to changes in the metal-cutting variables.

In order to describe the term "ploughing" in the cutting process it is necessary to understand what happens to the small portion of workpiece material in front of the rounded portion of the extreme edge, shown in Fig. 7. As the tool progresses in the cut, this metal in front of the cutting edge should move out one way or another. If this metal were to be pressed into the workpiece the tool would be constantly moved up and progress along some inclined line by the virtue of the fact that more material would be beneath the tool. Because experience shows that the tool does not move out of the cut, but keeps a constant depth, it may be concluded that essentially all material above the horizontal dashed line in Fig. 7 must be removed in the form of a chip. This includes the small portion of the metal in front of the cutting edge, which is pressed into and becomes part of the chip. This whole action will be termed here as ploughing.

Briefly, as the chip slides along the tool face, the metal in front of the rounded portion of the extreme cutting edge will be ploughed and pressed into the chip surface, as illustrated in Fig. 7. The plastic deformation of the metal in ploughing appears to be even more severe than that occurring on the shear plane. Also, ploughing usually occurs on a smaller scale. Generally, all the metal separated in the form of a chip has been sheared but only part of it along the chip surface has been ploughed.

The amount of material sheared depends essentially on the uncut chip thickness t_1 and the amount of ploughed metal depends on the magnitude of the sharpness radius r of the cutting tool, or on the radius of the built-up edge r_e (effective radius). Thus the ratio of these two dimensions, $\lambda = (r/t_1)$, will indicate in which proportion the shearing and ploughing take place in the cutting process.

In Fig. 7 it is shown that a small portion of the ploughed metal will be pressed into the work surface. Accommodating more material in the surface layer will cause compressive stresses, known as residual stresses in machined surfaces. These residual stresses may cause the workpiece to bend a little if its rigidity is not extremely high. As an analog to this event we can see that the metal pressed into the chip surface will cause similar compressive stresses in its surface and make the chip bend (so-called curling of the chip). Because much more of the ploughed metal will be pressed into the chip than into the workpiece, since the chip is usually much less rigid than the work piece, we have much more bending (curling) of the chip than of the workpiece. Alternately, it may be said, in describing these events, that the chip being flexible can more easily yield to accommodate more additional material in its surface.

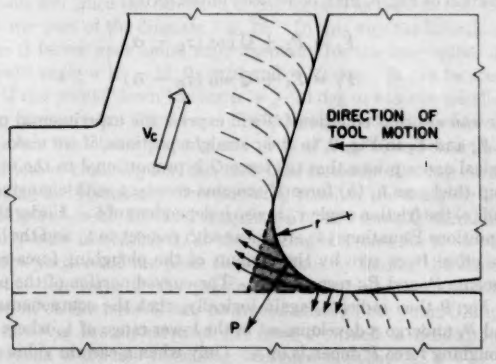


Fig. 7 Ploughing process. The shadowed area of metal in front of the sharpness rounding is displaced mainly into the chip; this flow of metal is indicated by arrows. The metal being pressed into the chip tends to expand the layer adjacent to the tool face, contributing in this way, to chip curling. A minor portion of the metal will be pressed into the newly produced work surface causing residual compressive stresses usually found after a machining operation.

New Force Diagram. By taking the ploughing process into account the tool force diagram will be significantly changed. If only the shearing process is accounted for, the tool-force diagram has a familiar look, Fig. 1, Ref. [1],[1] where only one resultant force R on the tool face can be resolved in the direction of the tool face and perpendicular to it, yielding the components F and N, or the resultant R can be resolved in the direction of the shear plane and perpendicular to its yielding the components F_s and F_n. Finally, another pair of components is of interest, that is the cutting force F_c and the thrust force F_t, which will be obtained by resolving the same resultant R in the direction of motion of the tool and perpendicular to it. This force diagram was derived by Merchant [1] and is used for the derivation of expressions for such quantities as coefficient of friction on the tool face, the shear angle relationship and others.

Referring now to Fig. 8, where rounding of the cutting edge of the tool is introduced, the summing up of forces along the straight portion of tool face between the points O and A give the force Q, inclined to the normal to the tool face by the friction angle τ_Q. The force Q also can be seen in the force diagram in the right part of Fig. 8. Proceeding further in summing up the infinitesimal forces from point A along the rounded portion of the edge toward point B it may be seen that each infinitesimal component changes its direction on account of the curved nature of the surface. Thus the resultant in the force diagram would be a polygon of infinitesimal forces in the form of an arc which ends at point B. We see that instead of the resultant R we have obtained the force Q plus an arc-shaped polygon.

It can be seen readily that in order to express the coefficient of friction on the tool face analytically one would not resolve into components the resultant R but only the force Q which would be resolved into the component Q_f along the tool face and component Q_n perpendicular to it.

The following derivation of the analytical expressions for the familiar quantities of the metal-cutting process will be presented on the basis of the more developed force diagram given in this section. The mathematical expressions thus obtained present a part of the improved analytical model of the metal-cutting process which will be tested against experiment, as described later in the paper.

A shear-angle relationship derived on the basis of the improved initial assumptions is presently under development and will be presented later in a subsequent paper.

[1] Numbers in brackets designate References at end of paper.

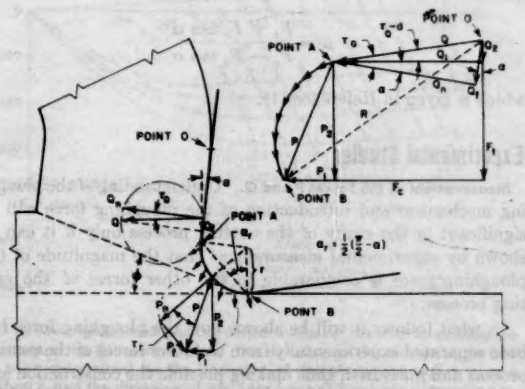


Fig. 8 New force diagram containing the force P which is due to the ploughing process

Components of the Forces P and Q. As was already mentioned, the arc polygon has the resultant P which can now be resolved into the normal force P_n and frictional force P_f, finding the direction for P_n by bisecting the angle between the direction of tool motion and direction of chip motion along the tool face. The point of location for P_n is the center of the circular segment representing the sharpness rounding. The direction of P_f is perpendicular to P_n. The angle between P_n and P can be designated as a friction angle τ_r. Denoting the direction of the tool motion by the index 1 and the direction perpendicular to it by the index 2 the other components of force P are described as P_1 and P_2.

Force Q can be resolved in a very similar manner into two components Q_n and Q_f which are the normal and frictional forces on the tool face, respectively, or it can be resolved into Q_1 and Q_2, components in the direction of the tool motion and perpendicular to it, respectively. The friction angle between the force Q_n and Q is denoted by τ_Q. For all these components the circle diagram representation as shown in Ref. [1] can be used, the only difference being that instead of one circle for R [1] there would be two circles, one for the force P and another for Q.

It can be seen from Fig. 8 that the cutting force F_c can be written in terms of components of P and Q as follows:

$$F_c = P_1 + Q_1 \tag{1}$$

and similarly the thrust force F_t:

$$F_t = P_2 + Q_2 \tag{2}$$

Coefficient of Friction on Tool Face. The considerations about the force components of Fig. 8 are sufficient for the derivation of an expression for the coefficient of friction on the tool face. According to the definition of the coefficient of friction:

$$\mu_Q = \tan \tau_Q = \frac{Q_f}{Q_n}$$

and expressing Q_f and Q_n in terms of Q_1 and Q_2, Fig. 8, we obtain

$$\mu_Q = \frac{Q_1 \sin \alpha + Q_2 \cos \alpha}{Q_1 \cos \alpha - Q_2 \sin \alpha} = \frac{Q_2 + Q_1 \tan \alpha}{Q_1 - Q_2 \tan \alpha}$$

Finally, using equations (1) and (2) we arrive at the expression for μ_Q:

$$\mu_Q = \frac{(F_t - P_2) + (F_c - P_1) \tan \alpha}{(F_c - P_1) - (F_t - P_2) \tan \alpha} \tag{3}$$

It can be seen that if the ploughing force P is neglected expression (3) reduces to the well-known expression for the coefficient of friction:

$$\mu = \frac{F_t + F_c \tan \alpha}{F_c - F_t \tan \alpha}$$

which is given in Reference [1].

Experimental Studies

Measurement of the Forces P and Q. Understanding of the ploughing mechanism and introduction of the ploughing force will be significant in the study of the cutting process only if it can be shown by experimental measurement that the magnitude of the ploughing force is comparable to the other forces of the cutting process.

In what follows it will be shown how the ploughing force has been separated experimentally from the other forces of the cutting process and measured, thus making possible the construction and confirmation of the new force diagram of Fig. 8 on the basis of experimental data.

If the cutting force F_c and the thrust force F_t are measured in the usual way the position of the points O and B of the force diagram Fig. 8 are established experimentally. Now instead of connecting these two points and obtaining the resultant R, as has usually been done, it is necessary to determine experimentally the directions of the forces P and Q and construct the lines in these directions through points B and O, respectively, completing the force diagram of Fig. 8.

In order to be able to find the directions of the forces P and Q it is necessary to study the behavior of these forces with increasing undeformed chip thickness t_1. For this purpose the cutting force F_c and the thrust force F_t have been measured in orthogonal cutting on a lathe for increasing t_1. The results of these measurments are shown in Fig. 9 where F_c and F_t are plotted against t_1 for different rake angles. Here it may be seen that the plots of forces are curved in the region of the smaller t_1 and at a certain value of t_1, or more precisely at a certain ratio of r/t_1, the plots become practically straight lines. In the present case this straightening out occurs at the uncut chip thickness of about $t_1 = 0.010 - 0.012$ in.

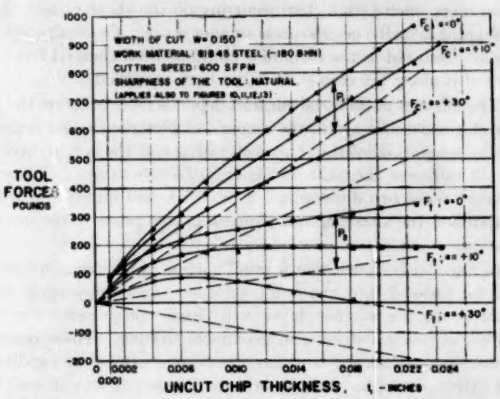

WIDTH OF CUT W= 0.150"
WORK MATERIAL: SIB45 STEEL (~180 BHN)
CUTTING SPEED: 600 SFPM
SHARPNESS OF THE TOOL: NATURAL
TOOL FORCES POUNDS
UNCUT CHIP THICKNESS, t, - INCHES

Fig. 9 Tool forces, dependent upon the uncut chip thickness t_1, illustrate the development of components P_1 and P_2 of the ploughing force P with increasing t_1 up to a certain stage when P remains practically constant with further increase of t_1

If we look now at the analytical expressions for the forces F_c and F_t, equations (1) and (2), which, by making use of the force diagram of Fig. 8, may be written in the form:

$$\left.\begin{aligned} F_c &= P_1 + Q \cos (\tau_Q - \alpha) \\ F_t &= P_2 + Q \sin (\tau_Q - \alpha) \end{aligned}\right\} \tag{4}$$

We can see that equation (4) will express the experimental plots of F_c and F_t in Fig. 9, in their straight portions, if we make the logical assumptions that (a) force Q is proportional to the uncut chip thickness t_1, (b) force P remains constant with changing t_1, and (c) the friction angle τ_Q is also independent of t_1. Under these conditions Equations (4) are linear with respect to t_1, and the lines are offset from zero by the amount of the ploughing force components P_1 and P_2, respectively. The curved portion of the plots of Fig. 9 then indicate, again logically, that the components P_1 and P_2 undergo a development in the lower range of t_1, where the ploughing force P depends on t_1. Only when a certain value of t_1 (or more specifically a certain ratio $\lambda = r/t_1$) is reached does the ploughing force P become practically constant with further increase of t_1.

Further conclusions which can be drawn from the results of the measurements of Fig. 9 are as follows: The curves for the cutting forces F_c increase monotonously with increasing t_1, but only one

thrust curve, for $\alpha = 0$ deg, is of the same character. The thrust force for $\alpha = +30$ deg increases with increasing t_1 only up to a certain maximum and then decreases and straightens out in a line having a negative slope, so that the thrust force decreases with increasing t_1 and finally becomes negative. The negative slope, as the expression for F_t, equation (4), shows, can occur if $\sin(\tau_Q - \alpha)$ is negative because of the friction angle τ_Q being smaller than the rake angle α. Since the rake angle was $\alpha = +30$ deg and the F_t curve has a negative slope in its straight portion we see that the friction angle on the tool face should be less than 30 deg. In fact these experiments show that if the rake angle is reduced from 30 deg, the negative slope of the F_t curve will decrease until it becomes parallel to the t_1 axis when α is of the order of 10 to 11 deg, as may be seen in Fig. 9. The zero slope of the F_t curve means that the friction angle in the experiments conducted was of the order of 10 to 11 deg. Decreasing the rake angle further we obtain a positive slope of the F_t plot, as may also be seen in Fig. 9 for $\alpha = 0$ deg.

Considering the force Q, represented by its components Q_1 and Q_2, we find that this force increases constantly with increasing undeformed chip thickness t_1, and never reaches a state of maturity where further development may be considered insignificant, as was the case for the ploughing force P. Such steady increase of Q with increasing t_1 is obviously due to the fact that the force Q is a reaction on the tool face due to the shearing forces and will therefore constantly increase with increasing t_1 as does S_s.

Keeping in mind all these considerations about the behavior of the forces P and Q one can proceed to the discussion of a special diagram which has been developed to permit the separation of the forces P and Q, or more precisely for the determination of directions of these forces. This diagram is shown in Figs. 10 and 11. Starting with Fig. 10 where the direction of the force Q has been determined, we see that this vector diagram relates the thrust force F_t to the cutting force F_c. The zero point of this diagram corresponds to the zero point of the force diagram of Fig. 8 and the point B of the same force diagram traces the curves shown in Fig. 10 when the uncut chip thickness t_1 is changed for a certain rake angle α.

It can be seen that these plots of F_t versus F_c are curved up to a certain uncut chip thickness t_1 and then become very nearly a straight line. The curved portions of the plots correspond to the phase where both forces Q and P are affected by the change in uncut chip thickness t_1. But as soon as the ploughing force P is fully developed and remains practically constant with increase in t_1 only the magnitude of Q increases further with increasing t_1. This means that with increasing t_1 the end point B of the force diagram will trace the direction of the force Q as may be seen in the lower part of the diagram Fig. 10. In this way the directions of the Q forces were found experimentally for the four values of the rake angle $\alpha = -10, 0, +10$ and $+30$ deg. It can be seen that Q can point "down" as for $\alpha = -10$ deg or can run parallel to F_c, or even point "up" ($\alpha = +30$ deg) depending on whether $\tau_Q \gtrless \alpha$. The linear parts of the plots of Fig. 10 indicate that the friction angle τ_Q on the tool face remains constant with changing t_1, because the direction of Q remains the same for increasing values of t_1. The magnitude of the force Q increases linearly with increasing t_1, as the spacing of points for t_1 on the straight portions of the plots shows. Thus by finding the directions of the force Q the problem is half solved because the direction of the force Q can be traced through the point O, Fig. 10; however the direction of the force P which is needed for completion of the new force diagram remains unknown.

Results of the experimental determination of the direction of force P are shown in Fig. 11. The direction of the ploughing force P can be determined by using the same vector diagram as used for the determination of the direction of the force Q. The only difference is that instead of t_1 another variable must be chosen, the variation of which leaves the force Q practically unaffected but changes the magnitude of the ploughing force P. Point B of the vector diagram, Figs. 10 and 11, will then trace the direction of force P. The variable which serves this purpose is the sharpness radius r. A change in r produces a change in the area of the surface of the rounded cutting edge, but since the character of the pressure distribution remains essentially the same, a change of magnitude of force P is obtained without significantly disturbing its direction.

The second requirement for the variable r, namely, that it should not significantly affect the force Q, can be seen to be realized to a high degree from the following considerations. For an idealized cutting process where ploughing would be eliminated only shearing remains and only the force Q will be present. Both forces being essentially due to different processes they may be varied practically independently by a suitable variable of each process. The small amount of interaction beweeen both forces, which occurs because the arc-shaped polygon has a small component P_s in the direction of the shear force: $S_s = P_s + Q_s$, can be neglected for our purposes, especially at the higher values of uncut chip thickness t_1. For these reasons the variable r has been used for the determination of the direction of the force P in the vector diagram, Fig. 11. Increasing the sharpness radius r by making use of a diamond lap, a set of four tools was prepared with size of radii as follows: r_0 (natural sharpness) = 0.0003 in., $r_1 = 0.0015$ in., $r_2 = 0.0032$ in., $r_3 = 0.0048$ in. Forces F_c and F_t

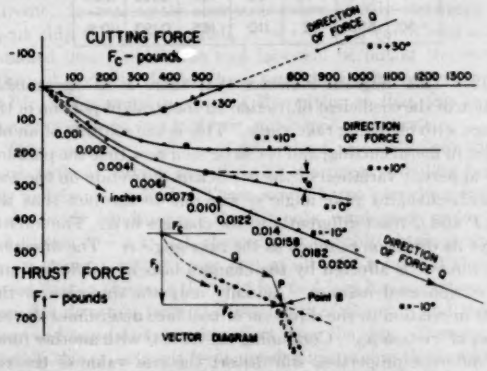


Fig. 10 Plot of F_t versus F_c which serves for experimental determination of the new force diagram. In this particular diagram the direction of the force Q on tool face has been determined for different rake angles by varying the uncut chip thickness t_1 in the range from 0.001 to 0.0202 in.

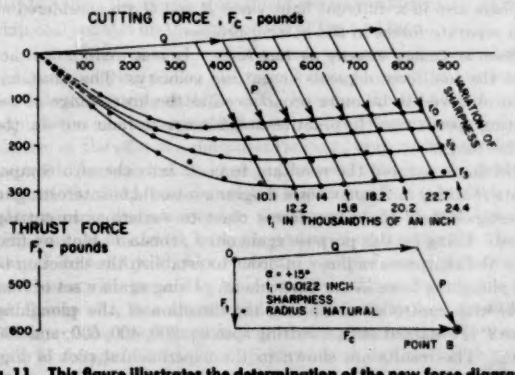


Fig. 11 This figure illustrates the determination of the new force diagram by means of an F_t versus F_c plot. The direction of the ploughing force P has been determined by the variation of the sharpness radius r, at a constant t_1 and the direction of tool face force Q has been determined by varying of t_1 keeping the sharpness radius r constant.

were measured on these four tools over a range of values of uncut chip thickness t_1 and plotted in the diagram, Fig. 11. By connecting the points for the same r and different t_1 the direction of force Q was obtained, and by connecting the points for the different values of radius r and the same t_1, the direction of force P was then found. In this way a family of new force diagrams, Fig. 11, has been obtained experimentally for various sharpness radii r and uncut chip thicknesses t_1. One of these new force diagrams is shown separately in the lower part of Fig. 11.

Determination of Coefficient of Friction on Tool Face. Having experimentally determined the force diagram in Fig. 11, it is possible to use these data for the calculation of the coefficient of friction on the tool face for a rake angle of $\alpha = +15$ deg. Similar force diagrams have been constructed from experimental data for the values of $\alpha = -15, 0, +30$ deg. The data from those diagrams used for computation of the coefficient of friction μ_Q according to equation (3) are shown in Table 1. The coefficients of friction μ_Q obtained are plotted against the rake angle α in

Table 1 Coefficient of friction on tool face when varying the rake angle α in the range of 45 deg

$\alpha°$	F_c LB	F_t LB	P_1 LB	P_2 LB	μ_Q	$\tau_Q°$
-15	895	520	25	95	0.196	11.1
0	805	310	65	170	0.190	10.75
+15	665	140	90	180	0.195	11.05
+30	555	12	110	165	0.193	10.9

Fig. 12. This diagram indicates that there is no appreciable variation of the coefficient of friction on the straight portion of the tool face with changing rake angle. This is the solution of an old paradox in metal cutting, and it can be seen now that the paradox of the apparent variation of the coefficient of friction on the tool face with changing rake angle α was due to the fact that the forces P and Q react differently to the changes in α. The force Q changes its direction as much as the rake angle α. The direction of the force P is affected by the changes in α in a different and more complicated manner. Actually only the direction of the force Q in relation to the direction of tool face determines the coefficient of friction μ_Q Combining the force Q with another force P, of different properties, will distort the real value of the coefficient of friction and endow it with peculiar properties such as dependence on the value of the rake angle.

The following section describes a mechanism by which the tool force is affected by the variation of cutting velocity. This effect appears also in a different light when P and Q are considered as two separate forces in the cutting process.

Effect of Cutting Velocity on Tool Force. It is a well-known fact that the tool force depends on cutting velocity. The effect has been observed to be more pronounced in the lower range of the cutting speeds used in practice and almost to fade out in the higher range.

Having separated the resultant force R into the two components P and Q in the new force diagram, it would be interesting to investigate how these two forces react to variations in cutting speed. Using for this purpose again our F_t versus F_c plot, we first vary the sharpness radius r in order to establish the direction of the ploughing force in our experiment. Using again a set of four tools with controlled sharpness, the direction of the ploughing force P is obtained at four cutting speeds; 200, 400, 600, and 800 sfpm. The results are shown in the experimental plot of Fig. 13(*a*) where it can be seen that the points obtained by the variation of the sharpness radius r (four points of the same marking represent different sharpness radii), for each of the various cutting speeds, fall along the same line with the points obtained by variation of cutting speed. The result points out that only the ploughing force P is affected by the variation of cutting velocity. If the tool-life force Q had also been affected significantly by the change in cutting velocity, the end point B of the force diagram would trace a direction different from the direction of the force P on account of the changing force Q. Thus it can be seen that the effect of the cutting velocity upon the tool force occurred through its effect on the ploughing force P alone.

Now let's consider the mechanism which makes the ploughing force P react to the changes in velocity. Having in mind that force P occurs on the very point of the cutter which moves through metal being plastically deformed at a very high rate of strain, it may be expected that the force P would increase with increasing velocity in much the same manner as would the force on the front portion of a body moving through a viscous fluid in hydrodynamics. But from experience it is known that the forces in metal cutting increase with velocity only in the very low range of cutting speeds. Reaching a maximum at a certain velocity (usually not very high) the forces start to decrease and, finally, practically level off with further increases in velocity.

This velocity characteristic of the cutting forces differs greatly from the characteristics known in hydrodynamics, where the force increases monotonously through a range of velocities. This difference in behavior points to the fact that there is in metal cutting some additional mechanism present which does not occur in hydrodynamic experiments. The additional mechanism present in the metal-cutting process is the formation of the built-up edge. The built-up edge formation changes the dimension of that portion of the tool where the ploughing force P occurs.

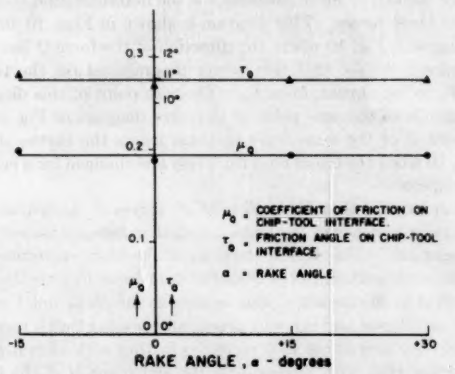


Fig. 12 Results of the experimental determination of the coefficient of friction on tool face for a wide range of rake angles

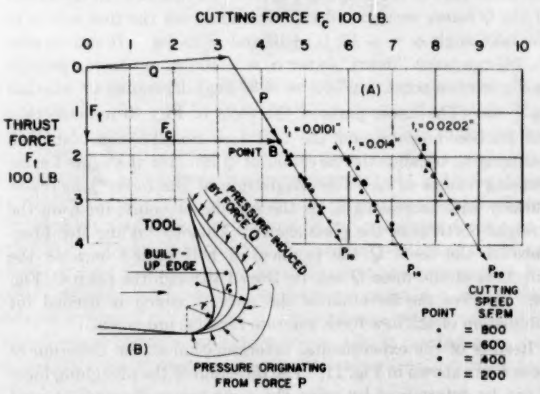


Fig. 13 The effect of the cutting velocity upon the tool force has been found to occur due to the development of the built-up edge. The four points of the same marking correspond to different sharpness radii.

Development of the built-up edge increases the rounding of the extreme cutting edge and hence the sharpness radius r. This causes the ploughing force P to change in much the same way as when the sharpness radius r is changed artificially. The occurrence of the built-up edge explains the fact that the direction traced by the points in Fig. 13(a) obtained by controlled sharpness coincides with the direction defined by points at varying cutting speeds, which also amounts to the changing of the radius r by variation in the size of the built-up edge. From the foregoing it may be seen that the decrease of cutting force with increasing cutting speed is due to the decrease of the ploughing force with diminishing of the size of the built-up edge in the region of higher cutting speed.

The foregoing considerations describe briefly the mechanism by which the cutting velocity affects the tool force. If we would like to represent this mechanism in terms of mathematical expressions or, in other words, to develop the mathematical model of the physical phenomena, we may start as follows: An effective sharpness radius r_e, Fig. 13(b), should be introduced which includes an increase of the sharpness radius r due to the presence of a built-up edge. The radius r_e consists of two components, $r_e = r + \Delta r$, one of which, the sharpness radius r of the cutting edge itself, is independent of cutting velocity v, and the other component Δr expresses the increase of r due to the presence of a built-up edge. Component Δr would depend on the cutting velocity and on some other variables. The significance of these are presently being investigated experimentally. It is intended that results of this investigation be presented in a future paper.

Results of Sharpness Measurement. All sharpness measurements in this research have been made by the author's method which was mentioned previously. Results of these measurements were used to provide the experimental points to establish the relation between sharpness radius r and included angle δ of the cutting edge. This relation, $r = r(\delta)$, has been discussed generally and was found to be of the type shown in Fig. 4 which is expressible by the formula $r = a \tan^n(\delta/2)$. As can be seen in the plot Fig. 14, the experimental points follow the trend of the curve suggested in general discussion of this relation. Because sharpness radius r for the edges of cutting tools has been found to be normally less than one thousandth of an inch, and often less than half a thousandth of an inch, it was found convenient, by the author, to express the sharpness in sharpness units instead of giving the length of the sharpness radius in inches. The sharpness unit (abbreviated "shu") has been defined here simply as one ten thousandth of an inch of the length of the sharpness radius r.

According to the experimental points of the plot, Fig. 14, the parameters a and n of the formula $r = a \tan^n(\delta/2)$, may be chosen to have values $a = 5.2$ and $n = 2$, so that the formula will be $r = 5.2 \tan^2(\delta/2)$ for the experimental conditions used.

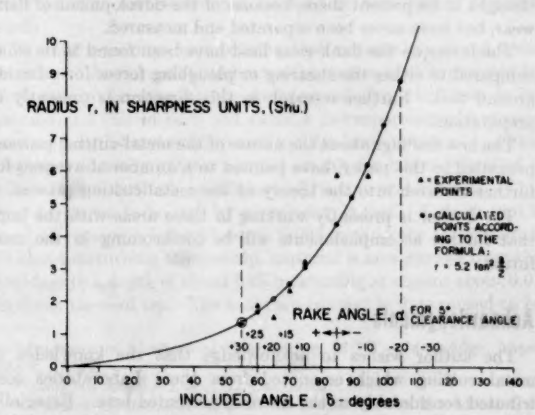

RADIUS r, IN SHARPNESS UNITS. (Shu.)
INCLUDED ANGLE, δ - degrees
▲ = EXPERIMENTAL POINTS
● = CALCULATED POINTS ACCORDING TO THE FORMULA: r = 5.2 tan² δ/2
RAKE ANGLE, α FOR 5° CLEARANCE ANGLE
+25 +15 +30 +20 +10 0 -10 -20 -30
0 10 20 30 40 50 60 70 80 90 100 110 120 130 140
0 1 2 3 4 5 6 7 8 9 10

Fig. 14 Experimental results indicate the nature of the dependence of tool sharpness on the included angle of the cutting edge

It can be seen that there is no simple answer to the question, "How big is the sharpness radius of a cutting edge?" This question cannot be answered by giving a single value for the sharpness radius. The author's measurements show that it can be expected that the natural sharpness radius of a cutting tool will range between 0.0001 and 0.0010 in.

A study of the relation between the magnitude of the sharpness radius r and included angle δ is presented in this paper but there are, naturally, many more variables which can affect the natural magnitude of the sharpness radius.

About the artificial changes of sharpness radius it can be said that there are several means to increase the sharpness radius (controlled sharpness). Perhaps the simplest way of all is hand stoning.

Experiments Relating the Ploughing Process to Chip Curling. As was discussed earlier, there is a simple relation between the ploughing process and chip curling. This leads to the idea that by controlling the amount of ploughed metal the curling of the chip can be changed. In order to evaluate this effect, one might attempt to eliminate ploughing from the cutting process and observe how the curling of the chip will react to this change. This could be accomplished by decreasing the sharpness radius to such a small value that it becomes practically insignificant. However, because experimentally this would be rather difficult, we might consider what would happen if we turn to the other extreme. Increasing the sharpness radius greatly and keeping the uncut chip thickness t_1 small, $\lambda = r/t_1 > 1$, conditions may be obtained under which the tool face will be out of the range of chip-tool contact. Under these conditions chip-tool contact will end in the rounded portion of the cutter so that this rounded portion will serve as a tool face under such conditions.

This rounded "tool face" will blend smoothly into the tool flank without having any "small" sharpness rounding between the tool face and flank. There would be no ploughing involved under such conditions. Experiments conducted under these conditions, $\lambda = r/t_1 > 1$, showed that the chip obtained was perfectly straight without any curl at all. Of course, it should be noted that in this case the rounded tool face has curvature in the opposite direction from the usual curl of the chip and this may have had some straightening effect upon the chip. However, continuing this experiment for increasing uncut chip thickness t_1, i.e., for decreasing λ, the curling started to appear again at a value of uncut chip thickness such that the rounded part of the surface ceased to function as a rounded tool face and started to act as a sharpness rounding producing a ploughing effect, or, in other words, when the uncut chip thickness was sufficient to provide chip-tool contact on the flat portion of the tool face.

Further experiments on chip curling were carried out with continuously varying t_1, i.e., the ratio λ was decreased continuously for one chip during a single cut. Because the amount of metal ploughed remains essentially constant, due to comparatively little change of the effective sharpness radius r_e, chip curling can be expected to become less and less pronounced with increasing thickness of the chip because of its higher rigidity. In fact, this phenomenon was observed in the experiment. A spiral shaped chip was obtained with curl of a greater curvature for the thinner chip followed by a chip of continuously decreasing radius of curvature with increasing t_1. In additional experiments the rate of change of t_1 was altered and this was reflected in the rate of change of the spiral curvature in the way expected. For the low rate of change of t_1 a tighter spiral resulted, and for a high rate of change of t_1 a loose spiral was obtained.

All of these preliminary experiments have shown quantitative agreement with the described relation between the ploughing process and chip curling. Further, the experimentation points to

the fact that the ratio $\lambda = r_n/t_1$ should have a pronounced role among the variables affecting chip curling in the analytical (quantitative) description of this event.

Complete Force Diagram of Orthogonal Cutting Process. It should be noted that, apart from the ploughing force P which occurs because of the ploughing process, there is still another force in the cutting process, that which acts on the wear land of the tool flank. We denote this force by the letter L (land force).

Furthermore, a small part of the ploughed metal adjacent to wear land is pressed into the workpiece, Fig. 6, contrary to the main part of the ploughed metal which is pressed into the chip. It might be appropriate to denote the small force performing this action differently, say by P', because of the different direction of its frictional component P_f' as compared with the direction of the frictional component P_f of the main ploughing force P.

The complete force diagram, which includes forces L and P', is shown in Fig. 15. Unlike the ploughing force P which has a considerable magnitude even for a freshly ground tool, the force L is small when there is no appreciable wear land on the tool flank.

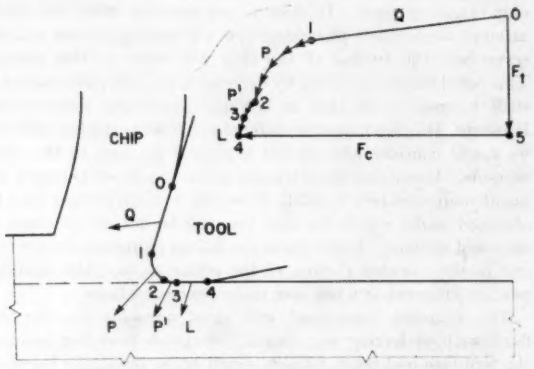


Fig. 15 The complete force diagram of the orthogonal cutting process

The numbering of Fig. 15 relates the forces shown in the force diagram to the portions of the engaged part of the tool where these forces occur. Thus between the points 0 and 1 of the tool face force Q is acting on the chip-tool contact area. Between the points 1 and 2 is the main part of the sharpness rounding (or front portion of the built-up edge if it is present). The pressures along this part of the engaged surface yields the arc-shaped portion of the force polygon between 1 and 2. This results in the ploughing force P. At point 2 the actual separation of the metal from the workpiece occurs. The motion of the separated metal occurs in the direction 2-1-0 along the engaged part of the cutter, and the metal of the workpiece slides in the direction of 2-3-4 along the tool flank. For this reason the friction forces P_f and P_f' oppose each other on opposite sides of point 2. The different direction of the friction forces results in a discontinuity of the force polygon at point 2. The angle formed by the two tangents to the force polygon at this point is equal to double the friction angle τ_r.

At points 1 and 3 discontinuity can occur if the coefficient of friction on the tool face and on the wear land differs from that on the sharpness rounding. If the coefficient of friction can be considered as constant along the whole tool surface over which metal slides, the portions of the force polygon from 0 to 2 and from 2 to 4 will be without discontinuities, as is shown in Fig. 15.

This complete force diagram was abbreviated in Fig. 8 where the force L was considered small enough to be neglected for the freshly ground tools with which the experiments were conducted. The force P' was also considered to be small and of much the same direction as the main ploughing force P. Thus in the force diagram of Fig. 8 the ploughing force P contains two small forces P' and L which, as we have seen, do not affect significantly the experimental results.

Conclusions

In this paper some of the results of recent research into the theory of the metal-cutting process have been presented.

Progress has been made mainly by recognizing the presence of the ploughing process in metal cutting. The mechanism of ploughing has been conceived and introduced into metal-cutting theory. This has made possible the construction of a more detailed force diagram which should serve as a basis for developing a more complete analytical model of the cutting process.

A method has been established for the experimental separation of the ploughing force P from other major cutting forces. Its magnitude has been found considerable as compared to the force due to shearing. The assessment of the tool-face force Q has allowed the determination of a real value of the coefficient of friction on the tool face μ_Q and has shown that the value of μ_Q does not vary with varying rake angle α. In this way the old paradox of the varying of the coefficient of friction on the tool face with rake angle α has been solved.

Relation of the ploughing process to the curling of the chip and to the residual stresses in the work surface has been established. Both events have been found to be essentially of the same nature, the difference being only a matter of degree, because of the difference in rigidity of the chip and of the workpiece. Bending of the workpiece due to the additional metal pressed into the machined surface due to ploughing is essentially the same event as bending of the chip (chip curling) due to the additional metal pressed into the chip surface by the ploughing process.

A study of the sharpness of cutting tools has been presented. The sharpness radius of cutting tools has been specifically measured and its influence on the metal-cutting process studied for the first time in metal-cutting research. These measurements have been carried out by the method of sharpness measurement which was devised by the author.

The magnitude of the sharpness radius was found to depend on several variables and usually to fall into the range between 0.0001 and 0.001 in. The higher values of this range were found to correspond to the negative range of rake angles and the smallest values, of a few tenths of a thousandth of an inch, were found to correspond to high positive rake angles.

Further, a complete new force diagram has been presented which contains, apart from newly found ploughing forces, the forces on the wear land of the tool flank, which have long been thought to be present there, because of the development of flank wear, but have never been separated and measured.

The forces on the flank wear land have been found to be small compared to either the shearing or ploughing forces for a freshly ground tool. Further research in this direction is presently in preparation.

The new findings about the nature of the metal-cutting process, presented in this paper, have pointed to a number of avenues for further research into the theory of the metal-cutting process.

The author is presently working in these areas with the hope that further accomplishments will be forthcoming in the near future.

Acknowledgments

The author wishes to acknowledge that the knowledge of metal cutting which emanated from these Laboratories contributed considerably to the research presented here. Especially the theory by Dr. M. E. Merchant, Director of Physical Research, presented in his paper "Mechanics of the Metal Cutting

Process," inspired the author to further study of the subject on the basis of the progress previously achieved.

The author is also very much indebted to Mr. E. J. Krabacher, Senior Research Supervisor, who contributed greatly in the preparation of this paper.

Appreciation should be expressed to the Laboratories' technical personnel whose skill was essential in performing the experimental work.

References

1 M. E. Merchant, "Mechanics of the Metal Cutting Process," *Journal of Applied Physics*, vol. 16, 1945, pp. 267–275.

2 Hans Ernst, "Physics of Metal Cutting," Machining of Metals, American Society for Metals, 1938, pp. 1–34.

DISCUSSION

Milton C. Shaw[2] and Nam P. Suh[3]

The idea that ploughing force components may be associated with the usual chip forming force components is not new as suggested by the authors' title. In 1953, Dr. Masami Masuko[4] put forth this thesis and developed it fully, including a discussion of shear plane implications. The treatment presented by Dr. Masuko is practically identical to that of the present author. Both analyses are based upon the following assumptions:

1 A rounded nose exists at the tool tip that is of significant size relative to the undeformed chip thickness t_1.

2 That the chip forming force Q is proportional to t_1.

3 That the coefficient of tool face friction is constant with t_1.

4 That the ploughing force P is independent of t_1 except at very small values.

5 That the normal component of ploughing force P bisects the included tool angle.

While the analyses of Dr. Masuko and the present author are consistent with the underlying assumptions, certain of these assumptions are questionable for the usual case.

First of all, practically all research work involving the analysis of cutting forces and stresses is conducted with *sharp tools* and values of undeformed chip thickness, that are 0.005 in. or greater. Since it is not difficult to obtain and maintain a cutting edge having a radius of 0.0002 in. or less at low cutting speeds, the size of the round nose represents an insignificant percentage of the depth of layer removed. While a rounded nose of the relative size shown in the author's Fig. 7 would undoubtedly have a significant influence on cutting force results, this would represent poor practice and would not be used in any serious research study.

It may be readily demonstrated that assumptions 2 and 3 are incorrect. As the author indicates the degree of chip curl will be observed to decrease as the undeformed chip thickness is increased and this in turn will cause a disproportionately greater contact length between chip and tool for large values of t_1. This disproportionately greater contact length for larger values of t_1 will influence both the coefficient of friction and force Q. This may be shown to be true by a very simple test. A sharp tool is used to remove a layer of depth 0.005 in. and the forces noted. Without disturbing the tool tip, material is now ground from the tool face to a depth of about 0.03 in. starting at a point about 0.01 in. from the tool tip. The tool-chip contact is thus caused to be less than the natural value by removing metal from the tool face. If the same cut is now repeated, the resultant cutting force and the coefficient of tool-face friction will be significantly altered. Yet, we have not altered either the tool-tip radius or the value of t. This can only mean that tool-chip contact has a significant influence on force Q. Thus when t_1 is varied and the tool-chip contact varies disproportionately due to the nonlinear increase in chip rigidity with increased thickness, we should not assume that Q varies linearly with t.

It would thus appear that the concept of a ploughing component is a valid one when cutting with a large built-up edge or very dull tool (relative to the undeformed chip thickness). However, in the usual research investigation the radius of the tool tip will be small relative to t_1 and hence the ploughing effect will be negligible as a photomicrograph of a partially formed chip will clearly show. Furthermore, it does not appear that the proposed method of finding the ploughing force P, when it is significant, is valid, since the method is based upon a questionable assumption that force Q is proportional to t.

[2] Professor of Mechanical Engineering, M.I.T., Cambridge, Mass. Mem. ASME.

[3] Research Assistant, M.I.T., Cambridge, Mass.

[4] Masami Masuko, "Fundamental Researches on the Metal Cutting. 1 A New Analysis of Cutting Force," *Trans. Society of Mechanical Engineers (Japan)*, vol. 19, 1953, pp. 32–39.

Author's Closure

The author wishes to thank Prof. M. C. Shaw and his associates for their interest in this paper. Thanks are due also for the information regarding the work of Dr. Masami Masuko who seems to have followed the same direction in developing further the metal-cutting theory. We, here, were unaware of this work and are now obtaining a translation of Dr. Masuko's paper so that it may be compared with the recent research of this laboratory.

Starting with the over-all impression of the discussion, we can see that the discussers missed a very important point, namely, that the developments presented in the paper are based on an analysis confirmed by experimental evidence and not on mere assumptions.

For instance, under item number 1 of their discussion the discussers say that the author made an assumption about the size of the natural sharpness rounding relative to the magnitude of the uncut chip thickness t_1. No such assumption was made. Instead, the natural sharpness radius was measured by the author and its magnitude in inches stated, permitting direct comparison with the magnitude of the uncut chip thickness. Thus it can be seen that there was no need of any assumption regarding the magnitude of the natural sharpness rounding because its magnitude was already known from the author's measurements. Regarding Fig. 7 and other illustrations it should be mentioned here that the magnitude of the sharpness radii shown in these diagrams was increased for the sake of clarity of illustration.

More specifically, with regard to the size of the natural sharpness rounding (as presented in the paper), it may be found by measurement that the natural sharpness radius is of a magnitude between 0.0002 and 0.0006 in. The lower values of this range usually occur on tools with a higher rake angle (such as +20 deg) and values of the order of 0.0006 in. may occur in the range of negative rake angles. These higher natural values of sharpness radius can be artificially altered to smaller ones by applying such techniques as fine polishing or lapping of the tool face and flank. This may be done for the sake of experiment and the author agrees that such a value as 0.0002 in. can thus be obtained on tools of low rake angle. However, it should be borne in mind that this will be an artificially obtained situation which takes one away from the ordinary conditions of a usual cutting operation.

It is interesting to note that the discussers seem to approve an artificial decrease of the sharpness radius (apparently in an attempt to make it very small and insignificant as compared with t_1) but an increase of the sharpness radius for experimentation the discussers proclaim immediately as "poor practice." The author wishes simply to state here that either increasing or decreasing the natural value of the sharpness radius are equally useful and sound techniques for research on the ploughing process.

Actually, the decreased and increased values of the sharpness radius are two branches of one range of controlled sharpness and there is no reason whatsoever to accept one side of the range and reject the other side of the same range for experimentation.

Now let us consider what might cause the discussers to be inclined to such a one-sided preference. As early authors began development of the present theory of metal cutting, certain simplifying assumptions were made, as is usually the case when starting a study of a complex problem. Among these initial simplifications was one about the size and effect of the sharpness radius being small and insignificant compared to the dimensions of the cutting mechanism. Since that time probably every one who raised the question about the significance of the sharpness rounding to the cutting process has heard the phrase "small and insignificant." This is not surprising, because there was practically nothing known about this very point of the cutting edge. Unfortunately, for many years no study was undertaken to confirm or replace this simplifying assumption made by the early investigators. Thus this initial assumption has, with the years, become a customary way of thinking.

Backed by experimental evidence, the developments presented by the author in this paper break the customary way of thinking about the significance of the sharpness rounding to the cutting process.

Now, let us look at the statement "small and insignificant" in the light of research presented in the paper. First of all, "small" was replaced by the results of measurements. The sharpness rounding has been assessed quantitatively and it can be seen that its magnitude is not very great as compared with t_1 or other dimensions of the metal-cutting mechanism. However, how about its "insignificance?" We can see again from the experimental results presented in the paper that the force due to ploughing is significant. Its value may attain hundreds of pounds, equalling in magnitude the other forces of the metal cutting process, in the range of smaller values of t_1. Thus we have arrived at the picture of a comparatively small dimension of the cutting edge with a significant force on it. Is this a controversial result? The answer is that this is not the final picture of what has been found. At this stage the author would like to give a preview of some points which will be discussed in Part II of the paper. One point will be the occurrence of fast initial wear at the tool edge, which happens because of the engagement of the cutting edge in the cut. This wear usually produces a slight increase in the initial sharpness rounding of the cutting edge, so that the edge actually cuts with a slightly larger rounding than that which initially resulted from grinding. Usually, however, this increase of the sharpness rounding is small and does not change the picture very much.

Another factor of greater significance is the presence of a small built-up edge which is still present under conditions which are usually termed as "cutting without the built-up edge." When cutting steel at speeds as high as 600 fpm, or even higher, this small built-up edge was observed under the microscope. The occurrence of even a small built-up edge increases the effective sharpness radius significantly and thus the ploughing effect and ploughing force. As a matter of fact, the observed development of the ploughing force with increasing t_1, as shown in Fig. 9 in the paper, has been found to be largely due to the development of this small built-up edge with increasing t_1. These are a few of the further steps made since Part I of the paper was written. It appears at this stage that the whole study of the ploughing process is intimately connected with the study of the behavior of the built-up edge. The study of the built-up edge should be extended to include not only the large built-up edge occurring at lower cutting speeds, where it may be observed with the naked eye, but also in the region of higher speeds, where it can be observed only through a microscope.

Summarizing, we can see that the size of the natural sharpness rounding itself is comparatively small (as demonstrated in the paper by the results of its measurement), but its effect upon the cutting forces is significant, mainly due to the fact that the effective sharpness radius is increased by the presence of a small built-up edge. The size of the effective sharpness radius, which includes the effect of the small built-up edge, depends only slightly on the cutting speed (in the region of higher cutting speeds) and depends more on the uncut chip thickness t_1, up to a certain value of t_1. Finally, the magnitude of the effective sharpness radius depends also on the sharpness radius of the cutting edge itself. This may be expressed by the formula $r_e = r + \Delta r$ (given in the paper) where Δr is the effect of the built-up edge. The magnitude of Δr is affected greatly by the cutting speed in the region of low cutting speeds. This effect almost fades out in the region of the higher cutting speeds where only a small built-up edge occurs and where Δr depends more on t_1 in the region of development of the ploughing force. Finally, it may be said that as soon as the small built-up edge is taken into consideration there will be practically no region of cutting conditions under which a trace of a built-up edge cannot be found. This means that cutting with a built-up edge seems to be the usual condition, at least when cutting steel.

Let us return now to the objections of the discussers regarding points 2 and 3 (of the discussion), that is, regarding the force Q being proportional to t_1 and the coefficient of friction on the chip-tool interface being constant with t_1.

Actually, the discussers' experiment (involving artificial reduction of the length of contact of the chip on the tool face), instead of tending to disprove these two points, tends to *confirm* the validity of the author's entire approach. If, as the discussers assume, there is no ploughing force or ploughing effect acting at the cutting edge, then the cutting force and chip friction will be determined entirely by the simple contact between the chip and the flat tool face only. One cannot conceive how, in such a symmetrical situation, artificially decreasing the contact length on the tool face could have any effect on the direction of the single resultant force generated by the pressure of the chip against the flat tool face—and thus any effect on the observed coefficient of friction. However, the discussers (as have other people before them), observed a *change* in the magnitude and *direction* of the resultant force on the tool when they artificially decreased the chip-tool contact length. This can only mean, then, that the picture of the entire resultant force being associated with the contact on the flat tool face is too simple; that there must be a second force system contributing to the resultant force in addition to the tool face force—namely, a force system at the cutting edge—a ploughing force. Then, artificially decreasing the contact length on the tool face will change the relative magnitude of these *two* force systems—that on the tool face and that at the cutting edge—thus changing the observed magnitude and direction of the over-all resultant force, even though the direction of the tool face force and the ploughing force do not change. This change in magnitude and direction of the resultant force—i.e., change in cutting force and apparent coefficient of friction—predictable only when two force systems are postulated, is exactly what the discussers observed. Thus, this simple experiment confirms the existence of a ploughing force and is consistent with the picture of a tool face force Q which is proportional to t_1 but constant in direction (coefficient of tool face friction independent of t_1).

G. D. BREWER
Member of Technical Staff,
Space Technology Laboratories,
Ramo-Wooldridge Corporation
Los Angeles, Calif.

E. LEVIN[1]
The RAND Corporation,
Santa Monica, Calif.
Assoc. Mem. ASME

Structural Weight Approximation for a Bell-Nozzle Divergent Section

In the preliminary design of missile systems it is necessary to consider the weight of the individual components which comprise the missile. The usual procedure in weight approximation is to consider that an article is composed of small regular-shaped bits, each of whose weights is easily found and whose sum approximates the weight of the total article. When the component is such an irregular shape that a simple approximation cannot be made, or where the number of digital operations is excessive, the estimation of its weight becomes a tedious procedure. In this report, a method is derived for approximating the structural weight of such an article, namely, the divergent section of a bell-shaped nozzle for a rocket motor.

THE WEIGHT OF THE STRUCTURE of the bell-shaped divergent section of a rocket-motor nozzle can be approximated by use of the equation

$$W = \rho V = \rho \frac{p_0 r_t^3}{\sigma_t} V_0(\gamma, M_e, \alpha)$$

where values for $V_0(\gamma, M_e, \alpha)$ may be obtained from the universal curves presented in this report.

Discussion

Rocket-motor nozzles having bell-shaped divergent sections have been demonstrated to be more efficient than those having straight conical expansion shapes, particularly when large expansion ratios are employed. This report presents a method of approximating the structural weight of bell-nozzle divergent sections.

The theoretically-ideal expansion shape expands the gas to the ambient pressure and turns the gas flow so that at the exit it is parallel to the nozzle axis. Thus there is no divergent flow aft of the nozzle and all the available momentum is axially directed. However, the exit-end portion of a theoretically-ideal design approaches a cylindrical shape asymptotically and, as a result, is very long. Further, the nearly cylindrical exit end contributes very little to the thrust output. An appreciable length and weight saving can be realized at very little sacrifice in efficiency by cutting the exit section short. However, when the theoretically ideal nozzle shape is cut short, it is no longer necessarily an optimum design. Other configurations for the same length and expansion ratio may provide more thrust. Consequently, several methods of designing bell nozzles for practical application have been developed. These include the Foelsch method, the Rocketdyne "overexpansion and overcompression" concept, and the Guderley method. Unfortunately, none of these provides a simple analytical expression for the nozzle configuration.

For purposes of approximating the weight of the structure of proposed bell-nozzle designs, use of a parabolic equation to represent the geometric shape of the divergent section was suggested.[2] The parabola passes through the nozzle throat, diverges at an angle of 35 deg and gives the required area ratio at a length which is 75 per cent of the length of the divergent section of a 15-deg conical nozzle of the same expansion ratio, Fig. 1. The expression is

$$L - x = \frac{Ar^2}{r_t} + Br + Cr_t \qquad (1)$$

where

$$A = C + \text{ctn } 35$$

$$B = -2C - \text{ctn } 35$$

$$C = -\left\{\frac{\epsilon \text{ ctn } 35 - \left(\frac{0.75}{\tan 15} + \text{ctn } 35\right)\epsilon^{1/2} + \frac{0.75}{\tan 15}}{(\epsilon^{1/2} - 1)^2}\right\}$$

$$\epsilon = \text{expansion ratio} = \frac{\text{exit area}}{\text{throat area}}$$

and L is assumed to be $\dfrac{0.75 r_t}{\tan 15}(\epsilon^{1/2} - 1)$

It is emphasized that the parabolic approximation to the bell-shape divergent section used herein was originated solely for the purpose of weight approximation and should not be used for hardware-design purposes.

The following equations were provided[2] for calculating the metal wall thickness required for structural purposes.

[2] Personal communication from Dr. C. J. Wang, Space Technology Laboratories, Aeronautics Laboratory.

[3] Personal communication from Dr. Abner Kaplan, Space Technology Laboratories, Aeronautics Laboratory.

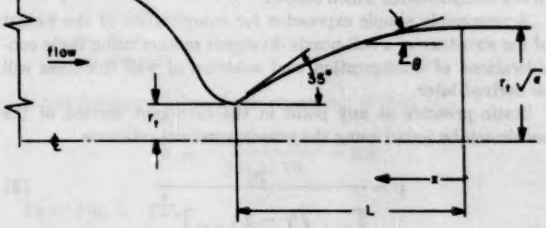


Fig. 1 Nozzle divergent section

[1] Formerly, Member of Technical Staff, Space Technology Laboratories.

Contributed by the Aviation Division and presented at the Annual Meeting, Atlantic City, N. J., November 29–December 4, 1959, of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS.

NOTE: Statements and opinions advanced in papers are to be understood as individual expressions of their authors and not those of the Society. Manuscript received at ASME Headquarters, August 15, 1958. Paper No. 59—A-69.

$$\tau_a = -\frac{P_x}{\sigma_t r}\left(1+\left(\frac{dr}{dx}\right)^2\right)^{1/2} \text{ (for axial stress)} \quad (2)$$

$$\tau_c = \frac{r}{\sigma_t}\left\{1+\left(\frac{dr}{dx}\right)^2\right\}^{1/2}\left\{p - \frac{\dfrac{d^2r}{dx^2}}{r\left[1+\left(\frac{dr}{dx}\right)^2\right]}P_x\right\}$$

(for circumferential stress) (3)

$$\tau_s = \frac{\tau_a - \tau_c}{2\left(\dfrac{\sigma_s}{\sigma_t}\right)} \text{ (for shear stress)} \quad (4)$$

where

$|\tau|$ = thickness measured normal to the wall
σ_t = allowable tensile yield stress
σ_s = allowable shear yield stress
r = radius to the nozzle wall measured normal to the axial center line
x = axial distance measured from the exit end of the nozzle
P_x = summation of the pressure forces in the x-direction acting over a radius $= \int_r^{e^{1/2}r_t} pr\,dr$
p = local static gage pressure

The axial and circumferential stresses are principal stresses. In general, over the divergent section of the nozzle, an element will be subjected to axial compression and circumferential tension. (The exception to this will occur near the exit end of an overexpanded nozzle.) In any case, it is easy to show that these principal stresses will be of opposite sign; hence it is necessary to consider shear. Naturally, the stress which requires the maximum thickness will design the structure at any given point.

The maximum-shear-stress theory of failure provides a conservative estimate. Taking the allowable shear yield stress σ_s to be one half the allowable tensile yield stress σ also results in a conservative estimate. The thickness of material required for structural purposes is computed from

$$\tau = \max\{|\tau_a|, |\tau_c|, |\tau_a - \tau_c|\}.$$

Since the axial and circumferential stresses are of opposite sign, the governing stress will in all cases be the shear stress; therefore the thickness required due to stress considerations is

$$\tau_s = |\tau_a - \tau_c|.$$

From a practical engineering viewpoint, the allowable thickness τ is not entirely determined by the stresses. Considerations such as fabrication and handling may place a limitation on the minimum thickness τ_m which may be tolerated. The expression for τ is therefore taken to be

$$\tau = \max\{\tau_s, \tau_m\}.$$

This implies that the wall thickness can vary along the length of the nozzle. Consideration of this possible variation is included in the computations which follow.

A reasonably simple expression for computation of the weight of the structure of a bell-nozzle divergent section using these considerations of configuration and selection of wall thickness will be derived later.

Static pressure at any point in the divergent section of the nozzle can be found using the one-dimensional relations

$$p = \frac{p_0}{\left[1+\left(\dfrac{\gamma-1}{2}\right)\mathbf{M}^2\right]^{\frac{\gamma}{\gamma-1}}} \quad (5)$$

and

$$\frac{r}{r_t} = \frac{1}{\mathbf{M}^{1/2}}\left[\frac{2}{\gamma+1}\left(1+\frac{\gamma-1}{2}\mathbf{M}^2\right)\right]^{\frac{\gamma+1}{4(\gamma-1)}} \quad (6)$$

where

p_0 = stagnation pressure
$\gamma = \dfrac{c_p}{c_v}$
$\mathbf{M}$ = local Mach number
r_t = nozzle throat radius

These relationships involve the basic assumptions of ideal gas, reversible adiabatic flow, and frozen equilibrium.

The volume of the structure of the nozzle-divergent section is given by expression

$$V = 2\pi\int_0^L \frac{r\tau\,dx}{\cos\theta} \quad (7)$$

where

τ = max $[\tau_s, \tau_m]$
θ = arc tan $\left(\dfrac{dr}{dx}\right)$ (see Fig. 1)

It is convenient to express the integrand in terms of Mach number $\mathbf{M}$. From Equations (5) and (6),

$$pr^2 = p_0(r_t)^2\left(\frac{2}{\gamma+1}\right)^{\frac{\gamma+1}{2(\gamma-1)}}\frac{1}{\mathbf{M}\left[1+\left(\dfrac{\gamma-1}{2}\right)\mathbf{M}^2\right]^{1/2}} \quad (8)$$

$$\frac{dr}{r} = \frac{(\mathbf{M}^2-1)d\mathbf{M}}{2\mathbf{M}\left[1+\left(\dfrac{\gamma-1}{2}\right)\mathbf{M}^2\right]} \quad (9)$$

Therefore

$$P_x = \int_r^{e^{1/2}r_t} pr^2\frac{dr}{r} = \frac{1}{2}p_0(r_t)^2\left(\frac{2}{\gamma+1}\right)^{\frac{\gamma+1}{2(\gamma-1)}}\int_{\mathbf{M}}^{\mathbf{M}_e}\frac{(\mathbf{M}^2-1)d\mathbf{M}}{\mathbf{M}^3\left[1+\left(\dfrac{\gamma-1}{2}\right)\mathbf{M}^2\right]^{3/2}}$$

or

$$P_x = \frac{1}{2}p_0(r_t)^2\left(\frac{2}{\gamma+1}\right)^{\frac{\gamma+1}{2(\gamma-1)}}\left[\frac{1+\gamma\mathbf{M}^2}{\mathbf{M}\left(1+\left(\dfrac{\gamma-1}{2}\right)\mathbf{M}^2\right)^{1/2}}\right]_{\mathbf{M}}^{\mathbf{M}_e} \quad (10)$$

From Equation (1)

$$\left\{\begin{array}{l} \dfrac{dr}{dx} = -\dfrac{1}{2A\left(\dfrac{r}{r_t}\right)+B} \\ \\ \dfrac{d^2r}{dx^2} = -\dfrac{2A/r_t}{\left[2A\left(\dfrac{r}{r_t}\right)+B\right]^3} \end{array}\right\} \quad (11)$$

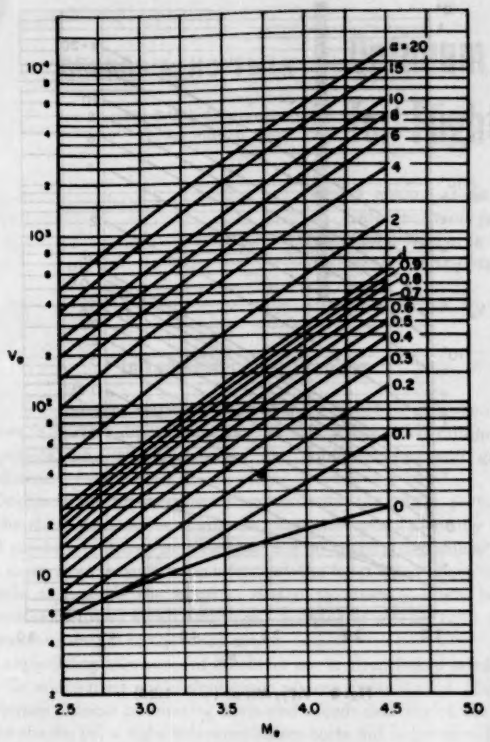


Fig. 2 $V_0(\gamma, M_e, \alpha)$ ($\gamma = 1.20$)

$$\tau_s = \left\{\left[1 + \frac{2Az}{G(G^2+1)}\right] H + \frac{2}{MF}\right\} \frac{p_0 r_i (G^2+1)^{1/2} \left(\frac{2}{\gamma+1}\right)^{\frac{\gamma+1}{2(\gamma-1)}}}{2\sigma_i G z} \quad (12)$$

where

$$F = \left(1 + \left(\frac{\gamma-1}{2}\right) M^2\right)^{1/2}$$

$$z = \frac{r}{r_i} = \left(\frac{2}{\gamma+1}\right)^{\frac{\gamma+1}{4(\gamma-1)}} \frac{(F)^{\frac{\gamma+1}{2(\gamma-1)}}}{M^{1/2}}$$

$$G = 2Az + B$$

$$H = \left[\frac{1+\gamma M^2}{MF}\right]_M^{M_e}$$

The volume can now be written

$$V = \pi(r_i)^2 \int_1^{M_e} z^2 (1+G^2)^{1/2} \tau \left[\frac{M^2-1}{MF^2}\right] dM. \quad (13)$$

The computation of V depends on six parameters: γ, r_i, σ_i, p_0, τ_m, M_e, where M_e = Mach number at the exit end of the nozzle. For any given set of values for these parameters, the volume can be computed from Equation (13). However, in order to obtain universal curves, it is convenient to introduce the dimensionless parameter

$$\alpha \equiv \frac{\sigma_i \tau_m}{p_0 r_i}.$$

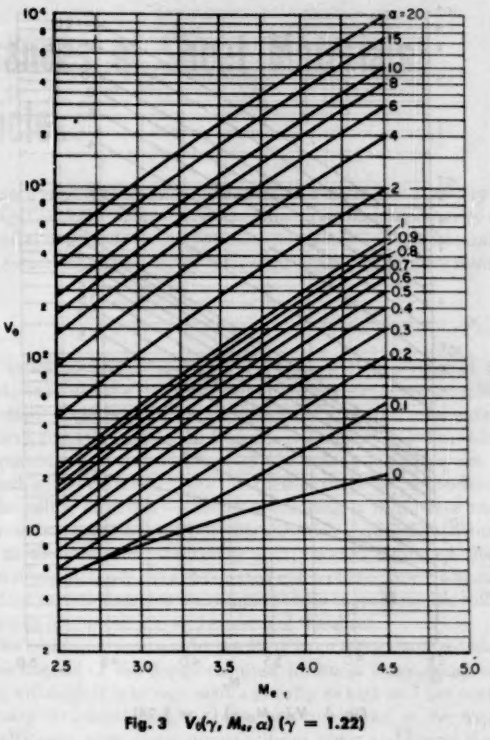


Fig. 3 $V_0(\gamma, M_e, \alpha)$ ($\gamma = 1.22$)

The volume can now be computed from

$$V = \frac{p_0 r_i^3}{\sigma_i} V_0(\gamma, M_e, \alpha), \quad (14)$$

where $V_0(\gamma, M_e, \alpha)$ is the value of V for $p_0 = r_i = \sigma_i = 1$. The quantity V_0 was obtained by numerical integration on the ERA 1103 computer for representative values of

γ = 1.20, 1.22, 1.24, . . ., 1.30
M_e = 2.5, 2.6, 2.7, . . ., 4.5
α = 0.0, 0.1, 0.2, 0.3, . . ., 1.0, 2.0, 4.0, . . ., 10.0, 15.0, 20.0

The results are presented as a set of universal curves in Figs. 2 through 7.

Sample Calculation

As an example of the use of this procedure, let it be required to determine the approximate weight of the structure of a nozzle-divergent section having the following characteristics:

γ = 1.22, M_e = 3.5 (this corresponds to an expansion area ratio of about 12.5)
σ_i = 20,000 psi
r_i = 10 in.
p_0 = 500 psi

If the minimum tolerable thickness τ_m = 0.05 in.,

$$\alpha = \frac{20{,}000 \times 0.05}{500 \times 10} = 0.2$$

From Fig. 3,

$$V_0 \cong 25.5.$$

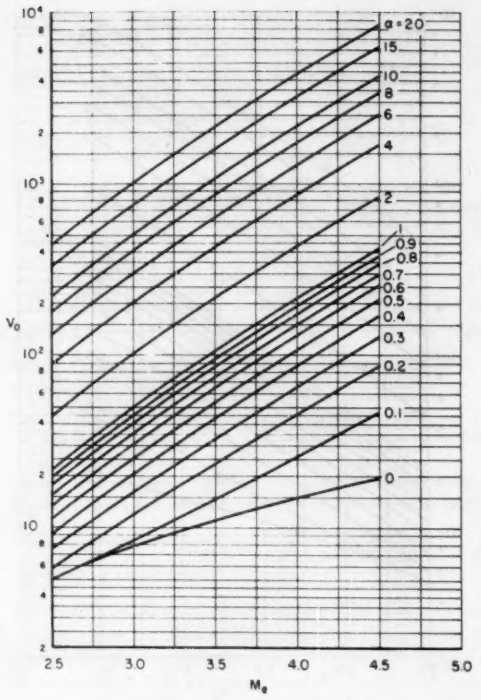


Fig. 4 $V_0(\gamma, M_e, \alpha)$ ($\gamma = 1.24$)

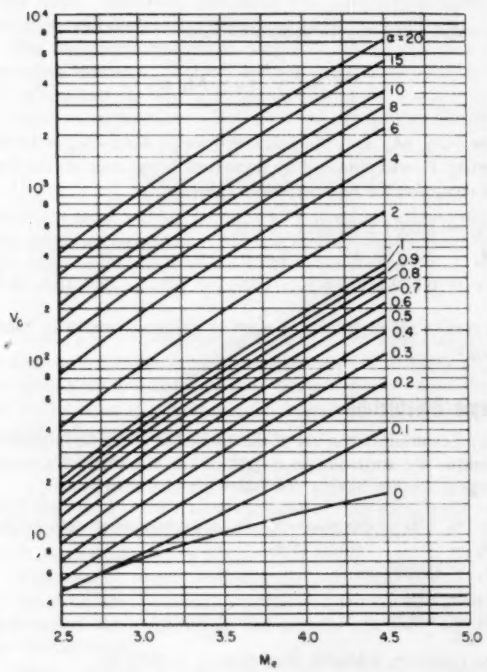


Fig. 5 $V_0(\gamma, M_e, \alpha)$ ($\gamma = 1.26$)

Hence, from Equation (13),

$$V = \frac{500 \times 10^3}{20{,}000} (25.5) \cong 640 \text{ in.}^3$$

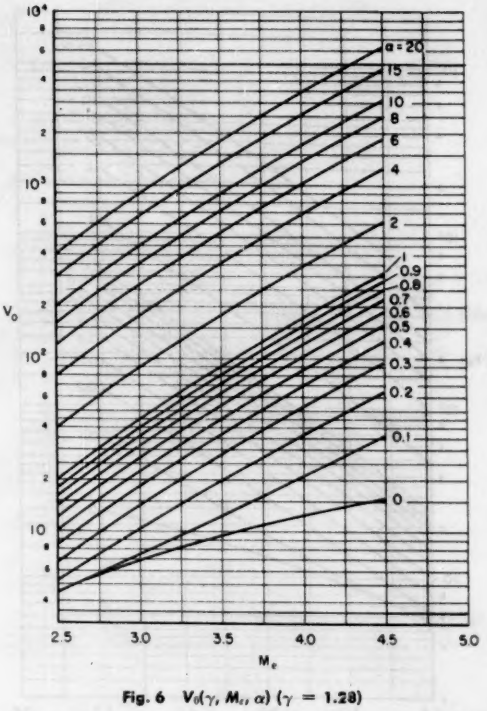


Fig. 6 $V_0(\gamma, M_e, \alpha)$ ($\gamma = 1.28$)

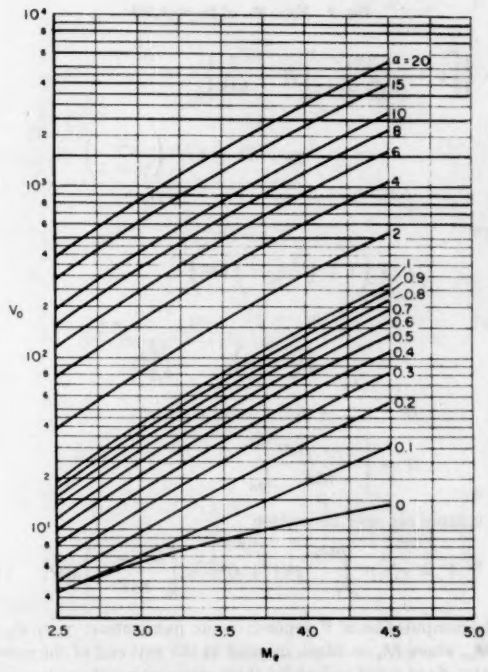


Fig. 7 $V_0(\gamma, M_e, \alpha)$ ($\gamma = 1.30$)

If the structural material is steel with density $\rho = 0.283$ lb/in ^{3}, the nozzle-divergent-section weight is

$$W = V\rho = 640 \times 0.283 \cong 180 \text{ lb.}$$

GEORGE A. HOFFMAN
Engineer,
The RAND Corporation,
Santa Monica, Calif.

Optimum Tolerances of Sheet Materials for Flight Vehicles

A method is derived for specifying economically optimum tolerances for a variety of sheet materials for use in flight vehicles. The results indicate a significant disparity between present industry practices and the apparent optimum tolerances. These tolerances are listed as possible target requirements for rolling programs of new and future metals.

Introduction: The Problem

SHEET DESIGN in flight-vehicle structures is governed by two factors—proper choice of a material with certain minimum properties, and specification of the configurations and gage thickness.

Designers usually assume, conservatively, that most portions of the sheet possess the minimum properties guaranteed by the mill producer. Likewise, they are apt to specify tolerances[1] on the sheet gage in such a manner that the least thick sheet used equals or exceeds the nominal design thickness. While both practices result in small additional margins of safety, both add extraneous weight to the structure.

In specifying tolerances of sheets for use in aircraft and missiles, one is confronted with conflicting alternatives: (*a*) A liberal tolerance reduces purchasing costs and speeds delivery of the desired sheets, (*b*) a tight tolerance raises costs but helps to reduce extraneous weight.

The present study develops a criterion by which the problem of tolerance specification can be solved. The method is based upon measuring both the extraneous weight and the additional cost as functions of the tolerance, in this way permitting a direct relation between the increased manufacturer's cost and the weight reduction from tightened tolerances. Identifying the manufacturer's cost with the worth to the vehicle operator of the weight saved determines a range of "most economical tolerances" for most materials and gages.

Another problem—not studied here, but bound to arise in connection with tightened tolerances—is selling this higher-priced product to the vehicle operator, that is, the problem of passing costs on to the consumer.

Extraneous Weight Imposed by Tolerances

The method of calculation used is applicable to most sizes, gages, and types of flight-vehicle sheet materials. For the sake of simplicity and as an illustration of possible applications within the titanium sheet-rolling program [1],[2] calculations are shown in this section only for titanium sheets with dimensions of 0.063 × 36 × 96 in. Results for aluminum, steel, and beryllium in a variety of gages are tabulated in a later section.

The extraneous material (i.e., the material beyond the thickness required by the design) in a manufactured sheet results from two types of deviation from the "perfectly" dimensioned flat sheet: (*a*) Variation in the thickness of the sheet along the rolling direction—experienced from sheet to sheet and, to a lesser extent, within a sheet; variations in the material properties at the rolling temperature and imperfections of the rolling machines are responsible; (*b*) variation in the thickness of the sheet perpendicular to the rolling direction—sometimes resulting in thin edges and a high center, called a "crown"; this variation is directly attributable to the design of the rolling mill; lately, improved sheet-fabricating methods have eliminated much of the crown, although random variations which are comparable to those along the rolling direction may occur across the width of the sheet.

The additional overweight resulting from using standard stock gages instead of the design-required thickness is not considered here, although it may represent a penalty as high as 7 per cent of primary structural weight. Also not considered is the weight penalty paid when the unavailability of wider and longer sheets requires the use of small sheets.

The overweight per square foot of sheet can be obtained from the formula

$$(0.5)^2 \times (\text{total magnitude of tolerance, in.}) \times (144) \times (\text{density})$$

0.5 = factor accounting for randomness of actual gages within the permissible tolerance of sheets used in fabrication and above the minimum design thickness. The factor is squared to account for randomness parallel and perpendicular to the rolling direction.

total magnitude of tolerance = maximum-minus-minimum thicknesses, measured in inches

$$144 = \frac{\text{sq in.}}{\text{sq ft}}$$

density =
0.067 lb per cu in. for beryllium
0.101 lb per cu in. for aluminum (7178)
0.162 lb per cu in. for titanium (6A1-4V)
0.276 lb per cu in. for steel (17-7 PH)

This formula was used for plotting Fig. 1.

Additional Cost Involved in Tightened Tolerances

Deviations from the required thickness of sheet can be held to a minimum by a number of methods. A few of these are listed in order of successively higher cost increments over commercial tolerance sheets:

1 Inspecting "all-over" sheets rolled to commercial tolerances (±0.006 for 0.063 gage), retaining, say, one half of the sheets with the best tolerances and reprocessing the remaining worse half. This procedure finds many a precedent in the large-scale scrapping for special reasons in aircraft production involving aluminum and titanium. By judicious inventory of these gaged-all-over sheets,

[1] Tolerances may be defined as the largest deviations from the nominal thickness of the sheet within some prescribed confidence level, usually between 0.90 and 0.95.

[2] Numbers in brackets designate References at end of paper.

Contributed by the Aviation Division and presented at the Annual Meeting, Atlantic City, N. J., November 29–December 4, 1959, of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS.

NOTE: Statements and opinions advanced in papers are to be understood as individual expressions of their authors and not those of the Society. Manuscript received at ASME Headquarters, April 17, 1959. Paper No. 59—A-70.

tolerance may be reduced to ±0.002 for 0.063 gage by selecting the best-dimensioned half of the sheets.

2 Rolling in special high-precision mills specifically designed to produce very accurate thicknesses. Such mills might yield 0.063-gage sheets with ±0.001 tolerance.

3 Sheet and plate skin grinding down to ±0.0005 in special mills (still in the design stage).

Approximate cost increments will be calculated for these three processes applied to a 0.063 × 36 × 96-in. titanium sheet. Costs are purposefully estimated on the high side and in 1959 dollars, and are to be interpreted only as an illustration of the method.

Reprocessing Sheets Above ±0.002 Tolerance. The incremental cost in dollars per square foot will be of the order of

$$(8.5 - 4.0) \times 0.063 \times 144 \times 0.162 = \$6.50 \text{ per sq ft}$$

where

8.5 = cost in dollars per lb of rolled sheet, Reference [2]; 1 lb is discarded as too thick for each lb used

4.0 = reuse value of discarded sheets = original value minus reprocessing costs such as gaging, grinding, polishing, and similar factors.

$0.063 \times 144 \times 0.162 = 1.47$ = weight in pounds of 1 sq ft of 0.063-in. titanium sheet

The cost of inspection, gaging, and handling should be well under $1.00 per sq ft; it was assumed to be negligible in comparison with the $6.50 per sq ft.

The cost of obtaining ±0.002 by reprocessing sheets happens to be about the same as the cost of skin rough-grinding down to ±0.002:

$$\frac{0.30}{0.001} \times (0.012 - 0.004 + 0.010) = \$5.50 \text{ per sq ft}$$

where

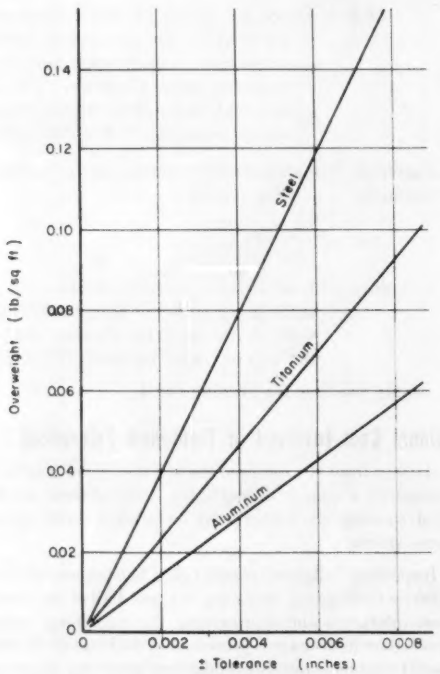


Fig. 1 Sheet overweights due to thickness tolerance

0.30 = cost of grinding 0.001 in. from 1 sq ft of titanium

0.012 − 0.004 = average thickness to be ground off = 2 × (tolerance before, minus tolerance after grinding)

0.010 = average thickness of available sheets minus desired average thickness

High-Precision Rolling to ±0.0010 Tolerance. A special mill—presently in use for other metals—would have to be utilized to roll titanium sheet to ±0.0010 tolerance. Such a mill, estimated to cost $6 million might be in use only half the time for titanium. The incremental cost in dollars per sq ft will be of the order of

$$\left(\frac{6{,}000{,}000 \times 1.2}{12} + 100{,}000\right) \times 0.5 \times \frac{1.47}{30{,}000} = \$17 \text{ per sq ft}$$

where

$6,000,000 = installed cost of mill, building, and so forth

1.2 = factor for upkeep, repair, replacement, utilities, and the like

12 = time period (years) of linear depreciation with zero resale value

$100,000 = cost to operate mill, per yr

0.5 = utilization factor for titanium

1.47 = conversion factor for sheet, lb per sq ft

30,000 = weight of titanium sheet that might be rolled in peak year, with industry pooling use of one mill. 30,000 lb per yr = 1/3 to 1/2 of total pounds in Table 2, Reference [1].

This cost of $17 per sq ft is high compared to the prospects of finish-grinding to ±0.001 as-delivered sheets. There are indications from present experiments that production grinding to such accuracy is feasible and might tentatively cost about $12 per sq ft for 0.063 sheet titanium.

Skin Grinding and Polishing to ±0.0005 Tolerance. Polishing mills capable of ±0.0005 production accuracy on titanium are yet to be designed. Tentative estimates have been made that the cost of achieving ±0.0005 may be double that of obtaining ±0.001.

These various costs are shown graphically in Fig. 2. The curves for aluminum and steel were obtained by similar calculations and from actual industrial experience.

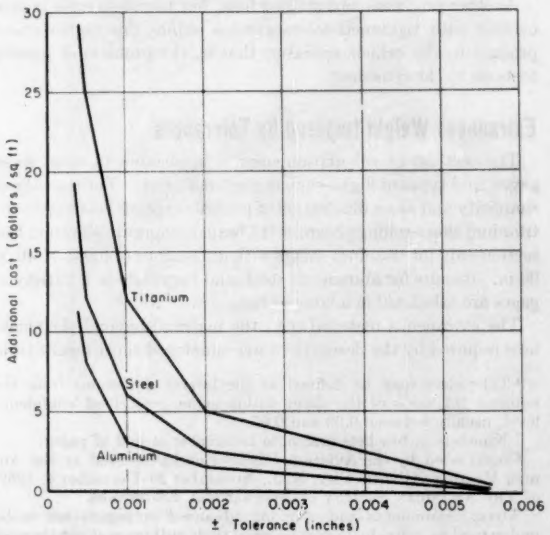


Fig. 2 The additional cost involved in tolerance control of 0.063 × 36 × 96-in. sheets

Criterion for Choosing Sheet Tolerances in Structural Materials

It has been shown, Figs. 1 and 2, that both cost and overweight can be functionally related by cost analysis to the same variable tolerance. It is now possible to relate the additional purchasing cost to the tolerance overweight in sheets, as in Fig. 3.

If an index or figure of merit were somehow devised for the value to an aircraft or missile operator of eliminating one pound of overweight, then it would be possible to read off directly from Fig. 3 the "optimum" tolerances on any thickness and type of sheet material.

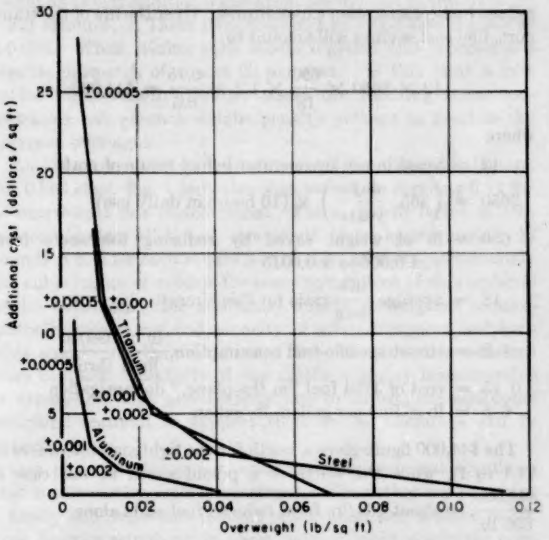


Fig. 3 The relation between additional purchasing costs and overweight in sheets

There is a great range of opinion as to the worth or value to the aircraft operator of a pound of weight saved. A list of some of these estimates and their sources may indicate the width of these divergent opinions: (*a*) The figure of $2065/lb has been mentioned for a commercial piston-engine transport [3]; (*b*) the figure of $600 to $800/lb has been derived for an American jet transport [4]; (*c*) correspondingly, figures of $560 to 840/lb have been used for a British jet transport [5], $220/lb for a fighter-interceptor [6], and $98/lb for a British propjet transport [5]; (*d*) space vehicles are most sensitive to extraneous weight—in actual experience it cost about $30,000 to place 1 lb of pay load into orbit in Explorer I, and about $15,000/lb for Explorer IV [7].

Because of this spread in estimated values, rays from $50 to $1000/lb are plotted in Fig. 4 for the worth of weight reduction in flight structures.

A criterion can therefore be formulated. For a wide range in the worth of a pound, there is only a narrow spread in the economically optimum tolerance. As an example, in the "worth" range of $200 to $500/lb saved, Fig. 4, aluminum and steel tolerances on 0.063 sheet can be specified to be ±0.001 in., while titanium tolerances appear to be between ±0.0015 and ±0.002. The accuracy of these economically optimum specifications is within ±50 per cent of their mean value in the range from $200 to $500/lb.

Discussion and Conclusions

It should be remembered that the plots in Fig. 4 and the desirable tolerances for steel and aluminum (±0.001) and titanium 0.063 sheet (±0.002) are sensitive to the annual production quantities of each material. This sensitivity becomes more apparent when one notices that the incremental cost of high-precision titanium sheet is inversely proportional to the lb/yr produced. Therefore, if the planned yearly poundages of titanium are revised significantly upward or downward, target tolerances should accordingly be revised, but in the opposite sense. The result is that, if the 30,000 lb per year used in figuring the additional cost involved in tightened tolerances is raised, for example, to 100,000 lb/yr, the optimum tolerance would be reduced to ±0.001. On the other hand, if the 30,000 lb/yr is lowered to 10,000 lb/yr, then the tolerance would go up to ±0.003.

One should also notice the trend for incremental cost to vary with gage thickness, all other factors remaining constant. The result is that, if the major portion of the titanium sheet produced is not in the 0.063-in. gage in Fig. 4, but is in the 0.032-in. gage, for example, then the optimum tolerance would be reduced to ±0.0015. There is a growing demand for sheet in increasingly thinner gages. In sandwich-type panels, the face sheets will be much less than half the gage of equivalent sheet-stringer designs; furthermore, thin titanium foil will be needed for the honeycomb or corrugated core.

To summarize these various possibilities, Table 1, the following assumptions were made: (*a*) The value to the vehicle operator of reducing 1 lb of overweight is $100/lb; (*b*) aluminum and steel sheet production is measured at approximately present-day levels; (*c*) titanium sheet may be produced at anywhere from 30,000 to 100,000 lb/yr; (*d*) production of beryllium sheet may be hypothesized at from 10,000 to 50,000 lb/yr; (*e*) tabular entries are shown for 36-in-wide sheets and are on the low side for much wider sheets.

The main conclusion of this study is apparent from Table 1; namely, that the economically optimum tolerances of all sheets of all materials are consistently smaller than those readily available at present in the aircraft and missile industry. It should be of major concern to realize that, in all instances, sheets used in flight vehicles today are overweighted by excessively loose tolerances.

Finally, this analysis serves to point out that expenditures constitute a design factor influencing structural weight in the manner of the more commonly studied factors, such as the configuration of

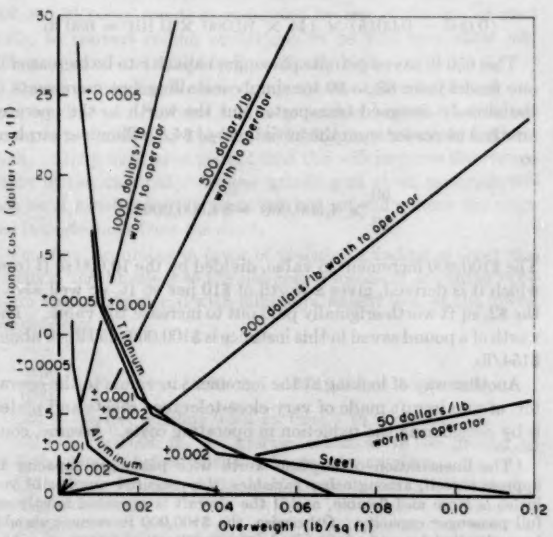


Fig. 4 Plots for choosing economically optimum sheet tolerances

Table 1 Optimum tolerances of sheet for various thicknesses and production levels

Sheet thickness	Beryllium 10,000 lb/yr (assumed)	Beryllium 50,000 lb/yr (assumed)	Titanium 30,000 lb/yr (assumed)	Titanium 100,000 lb/yr (assumed)	Aluminum at present production levels Optimum	Aluminum at present production levels Commercial	Steel at present production levels Optimum	Steel at present production levels Commercial
0.002	0.0003	0.0002	0.0002	0.0002	0.0002	0.0003	0.0002	...
0.020	0.0015	0.0012	0.0015	0.0010	0.0007	0.001	0.0008	0.002
0.063	0.0050	0.0030	0.0030	0.0020	0.0020	0.004	0.0020	0.006
0.125	0.008	0.005	0.006	0.004	0.003	0.005	0.004	0.010
0.250	0.012	0.008	0.009	0.006	0.005	0.013	0.005	0.017

structural elements and the material used [4]. For example, a sufficient expenditure increase when purchasing sheets could reduce structural weight by 3.6 to 5.3 per cent in a titanium-sheet structural component. (If the design thickness were 0.056 in., commercial sheet would have to be procured at 0.063-in. gage, ±0.006 in. The percentage of reducible overweight would then be $100 \times \frac{0.006 - 0.004}{0.056} = 3.6$ per cent, where 0.004 in. is a tolerance purchasable at some reasonable cost: while $100 \times \frac{0.006 - 0.003}{0.056} = 5.3$ per cent, where 0.003 is a tolerance purchasable at some moderately high cost.)

Structural weight reductions of this order merit much attention. They are accomplished without the least alteration in structural manufacturing practice, but are rather achieved by rearranging purchasing practices and convincing the vehicle operator of the advantages of such additional expenditures.

APPENDIX

Additional Examples of the Worth of a Pound of Material Saved

Consider the aluminum structure of a modern jet-transport aircraft such as the one described in Reference [8]. According to Table 1, tolerances between ±0.0020 and ±0.0007 in. should be specified, at additional costs up to $2/sq ft of surface. If this cost at first seems unduly high, the economics of the weight reduction may be illuminating. There are upward of 10,000 sq ft of sheet and plate in such a transport, with a weighted commercial (average) tolerance of ±0.006 in. Reduction to, say, ±0.0015 would save

$$(0.006 - 0.0015) \times 144 \times 10{,}000 \times 0.101 = 650 \text{ lb}$$

This 650 lb saved permits passenger capacity to be increased in one model from 88 to 90 by simply installing two more seats in the already designed transport. But the worth to the operator now has increased from the initial cost of $4.5 million per airplane to

$$\frac{90}{88} \times 4{,}500{,}000 = \$4{,}600{,}000^3$$

The $100,000 increment in value, divided by the 10,000 sq ft from which it is derived, gives a worth of $10 per sq ft, or well above the $2/sq ft worth originally paid out to increase the value. The worth of a pound saved in this instance is $100,000/650 lb, or about $154/lb.

Another way of looking at the increment in value, to the operator, of an aircraft made of very-close-tolerance sheets and plates is by considering the reduction in operating costs. Assume, conservatively, that the 650 lb saved is used by the operator only to reduce fuel-consumption expenditures. Over the life of the transport, fuel cost savings will amount to:

$$12 \times 3650 \times \frac{650}{18} \times 1.2 \times \frac{0.15}{6.5} = \$44{,}000$$

where

12 = years in use by operator before resale of craft

3650 = $\left(365, \frac{\text{days}}{\text{years}}\right) \times$ (10 hours in daily use)

650 = lb of weight saved by reducing tolerances from ±0.006 to ±0.0015

18 = average $\frac{\text{lift}}{\text{drag}}$ ratio for this aircraft

1.2 = thrust specific fuel consumption, $\frac{\text{lb of fuel/hr}}{\text{lb of thrust}}$

0.15 = cost of JP-4 fuel "in the plane," dollars/gallon

6.5 = lb of fuel per gallon, lb/gallon

The $44,000 figure gives a worth for the tightened tolerances of $4.4/sq ft, while the worth of a pound saved in this case is $\frac{\$44{,}000}{650 \text{ lb}}$, or about $68/lb, from reduced fuel costs alone.

References

1 N. E. Promisel and W. J. Harris, Jr., "The Titanium-Sheet-Rolling Program," *Mechanical Engineering*, vol. 79, 1957, pp. 1112–1115.

2 Titanium Review (Crucible Steel Company of America), vol. 6, August, 1958, p. 3.

3 H. Dreyfuss, "Designing for People," Simon and Schuster, Inc., New York, N. Y., 1955, p. 130 (quoting R. Gross of Lockheed Aircraft).

4 G. A. Hoffman, "Beryllium as an Aircraft Structural Material," *Aeronautical Engineering Review*, vol. 16, February, 1957, pp. 50–55, 82.

5 "The Financial Aspect of Weight Saving," *The Magazine of Magnesium*, November, 1957, pp. 10–11.

6 C. L. Baer, "Economics of Minimum Weight in the Weapon-System Era," Society of Aeronautical Weight Engineers, Paper No. 135, 1956, pp. 4–5.

7 Comments by Prof. H. J. Stewart at the Space Exploration Meeting, San Diego, Calif., August 5–6, 1958.

8 I. Stambler, "Boeing Designs 707 Structure for Long Fatigue Life," *Aviation Age*, August, 1957, pp. 26–31.

DISCUSSION

Lewis H. Abraham[4]

The author has pointed out a long neglected area for weight saving. Perhaps this inefficiency was tolerable in the past but with modern high performance aircraft, extreme range missiles, and spacecraft the argument becomes not one of can we afford to reduce tolerances but rather of can we afford not to reduce these weight penalties. When one considers the tremendous all out effort required to put just a few pounds out into space it is

[3] The linearization of airplane worth with passenger capacity is approximate if, among other variables, the procured number of vehicles is large and flexible, and if the aircraft is operated mostly at full passenger capacity. Otherwise, the $100,000 increment should be multiplied by some integrated future-utilization factor, possibly as low as 40 per cent.

[4] Chief, Strength Section, Douglas Aircraft Company, Inc., Santa Monica, Calif.

interesting to note that this total payload may be commensurate with the tolerance penalty contained in the boosting vehicles.

Fig. 1 does not include such materials as molybdenum, niobium, tungsten, or tantalum. If these materials were included this figure would be even more frightening.

The absolute value of the costs set forth in this paper may be debatable but the conclusions are none-the-less clear, perhaps even ominous. Rather than debate the costs I would be inclined to debate the time scale of feasibility.

While the paper has made an excellent presentation of the dollars and cents aspect of optimum tolerances there are many intangibles associated with the problem that indirectly can cause weight as well as additional costs that are difficult to tie down.

For example, in Table 1 the tolerance on 0.063 steel is given as ±0.006. When joining such sheets together this represents a possible mismatch of almost 20 per cent. If this joint is in a welded high-strength pressure vessel the resulting stress concentration can cause a weight penalty perhaps as great as the tolerance over again.

For the moment let us look at some of the other side factors. For 0.063 steel, Fig. 1 indicates that we will be carrying 0.12 lb/ft² overweight due to tolerances. For a growth factor of 100, and considering JP-4 fuel, this would mean approximately 12 pounds of fuel for each square foot of sheet area or approximately 425 cubic inches of volume for every square foot of this material used. Considering the additional missile power-plant requirements, increased size and capacity of ground support and handling equipment, the evaluation of these tolerances becomes even more complex. Similarly, if the missile is highly maneuverable or experiences high accelerations due to thrust, the additional structural material to support these excess tolerances can be important.

One phase of this paper, aside from its technical soundness that is noteworthy, is its succinctness. The author has presented a timely and important concept in less than four printed pages. It is likewise refreshing in these days of giant electronic computers to see a job done with simple arithmetic laced with sound engineering judgment.

Paul J. Queyrel[5]

The precision grinding of sheet materials for use in hypersonic vehicles is no longer in the development stage. It is available today and is being offered as a service by several firms. Consequently, precision grinding of sheet materials to gain precision tolerance is today's answer to today's problem.

The author has well stated the reason for precision grinding or acquiring precision sheet through other means and I will not take time to further prove that precision tolerances are necessary to gain weight reduction in high speed vehicles. I will, however, endeavor to show that precision grinding is the means by which close tolerance sheet should be developed.

Nearly all materials have been ground at one time or another. In our plant, we have ground magnesium, aluminum, all of the stainless alloys, many of the H-11 die steels, Invar, Inconel, molybdenum, and most of the titanium alloys. We have also worked on many of the chrome-molys used in rockets.

Precision sheet grinders are available that will handle work up to 7 ft wide by 40 ft long. The thickness of materials have ranged from 0.003 in. to several inches. Tolerances that have been achieved have been related to the gauge to which the material was ground, sheet size and type of material being ground. The tolerances to be anticipated should be in the neighborhood of plus or minus 0.0005 in. for materials up to 0.012 in. thick. From 0.012 to 0.140 in., tolerances of plus or minus 0.001 in. can be achieved. From 0.140 to 0.187 in., tolerances of plus or minus 0.0015 in. Above 0.187 in., wider tolerances should be expected. If weight is critical, tolerances of plus or minus 0.002 in. could be achieved on the heavier material.

The cost of obtaining close tolerance sheet is related to the hardness of the alloy, gauge to which it is to be ground, size, and quantity. The cost of grinding will run from 0.05 to 0.50 cents per square foot per thousandth of material to be removed. To use the same illustration that the author used, the cost of obtaining plus or minus 0.002 in. on 0.063 in. titanium sheet 36 × 120 in. was calculated by the author at 0.30 cents per square foot per thousandth, multiplied by the difference in the as-rolled condition of plus or minus 0.006 in. less the finished tolerance of plus or minus 0.002 in. plus 0.010 in. which is $5.50 per square foot.[6] I would like to correct his formula by eliminating the plus 0.010 in. which apparently was used to eliminate surface defects which may be involved but which is not to be considered as part of this problem. The author explained the 0.010 in. as being the average thickness of available sheets minus the desired average thickness. This certainly is not part of the cost of gaining precision material. It is the cost of reducing material an additional 0.010 in. beyond the available rolling of material. This is another problem and should not be considered in this formula. Consequently, my calculations show that the cost of precision grinding titanium sheet using the worst conditions would be a maximum of $2.40 per square foot.[7] This price could again be reduced by high level production to approximately $1.68 per square foot. By high level production I am only referring to 25 sheets per day. Within a few months, the cost of obtaining plus or minus 0.001 in. will be very little more than achieving plus or minus 0.002 in. today. The cost of grinding aluminum or magnesium is less than that of grinding titanium. The mild steels, 4130, 4340, H-11, etc., are ground for less than titanium. I wish to point out, however, that some of the stainless steels and high nickel content sheets will range in price close to, and in some instances above, that of titanium.

In many instances tapering, either simple or compound, can be accomplished while precision grinding the sheet material. The cost of developing tapers in sheet materials by grinding will run considerably less than any other method such as milling, either chemically or machine.

The use of precision sheet to the tolerances of plus or minus 0.001 in. as now being specified on many current pieces of hardware, is a custom requirement. It is difficult to imagine $400,000,000.00 being spent, as requested by one of the major steel mills, to convert rolling equipment to do this very same job. The per pound cost would be fantastically high.

Precision grinding to close tolerance also gains improved surface conditions. Surfaces can be ground as low as 6 microinch and can be finished by subsequent buffing operations to 3 microinch. Many tests have proved that this will improve the formability of the material. Proper grinding of sheet materials will not leave notch-sensitive areas but instead will remove the original imperfections from the sheet.

One final argument in favor of precision grinding of sheet materials is that it offers to the designer a means of acquiring optimum strength material with minimum weight loss, at the lowest possible cost.

---

[5] President, Mill Polishing Corporation, Huntington Park, Calif.

[6] $\frac{0.30}{0.001} \times (0.012 \text{ in.} - 0.004 \text{ in.} + 0.010 \text{ in.}) = \$5.50$ sq ft.

0.30 = Cost of grinding 0.001 in. from 1 sq ft of titanium.

0.012 in. − 0.004 in. = Average thickness to be ground off = 2 × tolerance before, minus tolerance after grinding.

0.010 in. = Average thickness of available sheets minus desired average thickness.

[7] $\frac{0.30}{0.001} \times (0.012 \text{ in.} - 0.004 \text{ in.}) = \$2.40$ sq ft.

Author's Closure

The author wishes to thank Messrs. Abraham and Queyrel for their constructive comments. As Mr. Abraham points out, the picture is indeed "more frightening" for the dense refractory metals. Using Mr. Queyrel's realistic and lesser costs shows that optimum tolerances of molybdenum sheet may be of the order of one tenth of the conventional tolerances. Mr. Abraham correctly points out that including other dollar-measurable factors (besides primary weight reduction) affects the results, leading to even more stringent economic tolerances than those listed in Table 1.

The author welcomes Mr. Queyrel's contribution emphasizing the well-developed techniques for achieving very tight sheet tolerances. It is encouraging to note how costs (stated in the paper to be purposefully high and, by now, outdated) have been reduced. This trend further lowers the optimum tolerances beyond the values in Table 1. For example, the downward revision of precision-grinding costs causes the optimum tolerance of 0.063 titanium sheet to be halved to ± 0.0010 to ± 0.0015, a still more ominous conclusion, to paraphrase Mr. Abraham.

Mechanical Vibration Filters With Nonlinear Characteristics

WILL J. WORLEY

Professor, Theoretical and Applied Mechanics, University of Illinois, Urbana, Ill. Mem. ASME

The behavior of a single degree of freedom system consisting of a single mass mounted on a spring and damper attached to an oscillating base is investigated. Steady-state and transient sinusoidal oscillations are applied to the base to which the suspension is attached. The response of the mass is recorded for various combinations of linear and nonlinear springs and dampers. Solutions are obtained with a differential analyzer.

Introduction

THE STEADY-STATE and transient solutions for the motion of a mass coupled through a linear spring and viscous damper to a vibrating base are well known. To determine whether a nonlinear spring and a nonlinear damper will improve the isolation characteristics of a suspension for steady state as well as for transient vibrations is the subject of this paper.

A linear spring and a spring having a hyperbolic tangent load-deflection characteristic were used in the analysis. The hyperbolic tangent spring characteristic has been studied by Mindlin [1][1] in connection with package cushioning. The hyperbolic tangent spring characteristic yields the "softening" type load-deflection behavior shown in Fig. 1.

The various types of damping were expressed in terms of the velocity of the mass relative to the velocity of the vibrating base. The general form of the expression for the damping force was $c|\dot{x}|^{n-1}\dot{x}$. The specific values of n were 0.5, 1, 2, and 3 where $n = 1$ indicates viscous damping while $n = 3$ would indicate a damping force which is proportional to the cube of the relative velocity. The force-relative velocity characteristic for various values of n appears in Fig. 2. Two other conditions, $n = 0.5$ for $\dot{x}$ negative and $n = 2$ for $\dot{x}$ positive as well as $n = 2$ for $\dot{x}$ negative and $n = 0.5$ for $\dot{x}$ positive were also used.

Rosenberg and Wang [2] investigated the steady-state solution for $n = 2$ when used with a linear spring. They found good agreement with the "equivalent viscous damping" concept proposed by Jacobsen [3]. This concept was used in the present paper to determine the equivalent coefficients for the various types of nonlinear damping.

Equation and Conditions Investigated

Consider a body which is mounted on a vibrating base subjected to sinusoidal oscillations of amplitude x_0. The spring force-deflection characteristic is assumed to be $P_0 \tanh (k_0 x/P_0)$ while the damper force-velocity characteristic is assumed to be $c|\dot{x}|^{n-1}\dot{x}$. Further it is assumed that the system oscillates in the gravitational field. This leads to the following equation of motion in terms of the displacement, velocity, and acceleration of the body with respect to the oscillating base:

$$\ddot{x} + \left(\frac{c}{m}\right)|\dot{x}|^{n-1}\dot{x} + \left(\frac{P_0}{m}\right)\tanh\left(\frac{k_0 x}{P_0}\right) = -g + a_0\omega^2 \cos \omega t \qquad (1)$$

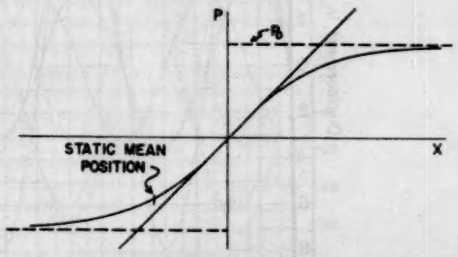


Fig. 1 Linear and hyperbolic tangent spring characteristics

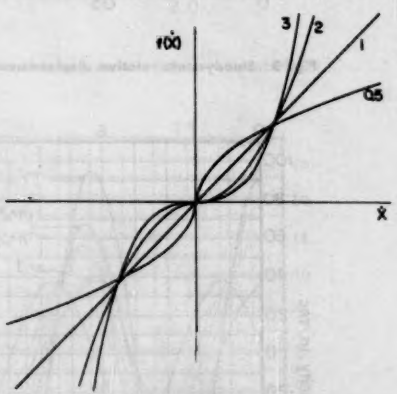


Fig. 2 Linear and nonlinear damping characteristics

[1] Numbers in brackets designate References at end of paper.

Contributed by the Machine Design Division and presented at the Annual Meeting, Atlantic City, N. J., November 29–December 4, 1959, of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS.

NOTE: Statements and opinions advanced in papers are to be understood as individual expressions of their authors and not those of the Society. Manuscript received at ASME Headquarters, December 24, 1958. Paper No. 59—A-42.

Nomenclature

m = mass, lb sec²/in.
g = acceleration of gravity, in./sec²
c = damping coefficient, lb sec^n/in.^{n}
P_0 = maximum hyperbolic tangent spring force, lb
k_0 = hyperbolic tangent spring constant at $x = 0$, lb/in.
ω = circular frequency of excitation, rad/sec
ω_n = natural frequency for linear spring-mass system, rad/sec
a_0 = excitation amplitude, in.
x_0 = excitation displacement, in.
x_1 = absolute displacement of the mass, in.
x = relative displacement of the mass, in.
n = exponent defining damping force, dimensionless

Since this equation was solved by means of a mechanical-electrical differential equation analyzer [4], it was necessary to assign specific constants to the system. These constants were $k_0 = 12$, $m = 1/12$, (weight 32.2 lb), $P_0 = 48$, and $a_0 = 0.4$, 0.8, 1.6, 2.0, 2.4, and 3.2. The damping coefficients c were selected for each value of n such that they would yield equivalent viscous damping at a frequency of $\omega = 12$ and $a_0 = 0.8$, corresponding to a damping ratio of 0.2 for the viscous damper. For $n = 2$, Jacobsen's procedure leads to $c = (3/8)(\pi c_1/\omega a_0)$ where c_1 is the coefficient for linear damping and a_0 is the amplitude of excitation of the base. The value of c_1 was 0.4. For $n = 0.5$, 2, and 3, the damping depended on both the amplitude and frequency of the relative displacement or since ωX represents the amplitude of $\dot{x}$, it could be stated that c depends on the amplitude of the relative velocity which in turn varies with ω and X. The particular choice of mass, with the given nonlinear spring characteristic resulted in the static mean position shown in Fig. 1, of -3.24 in.

Solutions at frequencies corresponding to $(\omega/\omega_n) = 0.5$, 1.0, and 2.0 were obtained in the form of relative displacement and relative velocity as a function of time.

Before proceeding to the discussion of the results it should be noted that Equation (1) may be expressed in dimensionless form as

$$\ddot{\sigma} + \frac{q_n}{r^2}\,|\dot{\sigma}|^{n-1}\,\dot{\sigma} + \frac{u}{r^2}\tanh\left(\frac{\sigma}{u}\right) = -\frac{h}{r^2} + \frac{1}{r^2}\cos\tau \quad (2)$$

where

$$\sigma = \frac{x}{a_0 r^2}, \qquad \dot{\sigma} = \frac{\dot{x}}{a_0 \omega_n r^3}, \qquad \tau = \omega t, \qquad r^2 = \frac{m\omega^2}{k_0},$$

$$q_n = \left(\frac{c a_0^{n-1} r^{3n-2}}{k^{1-(n/2)} m^{n/2}}\right), \qquad u = \frac{P_0}{a_0 k_0 r^2}, \qquad h = \frac{mg}{a_0 k_0 r^2}$$

Figs. 3 through 10 are expressed in both dimensional and in dimensionless units. The dimensionless values appear at the top

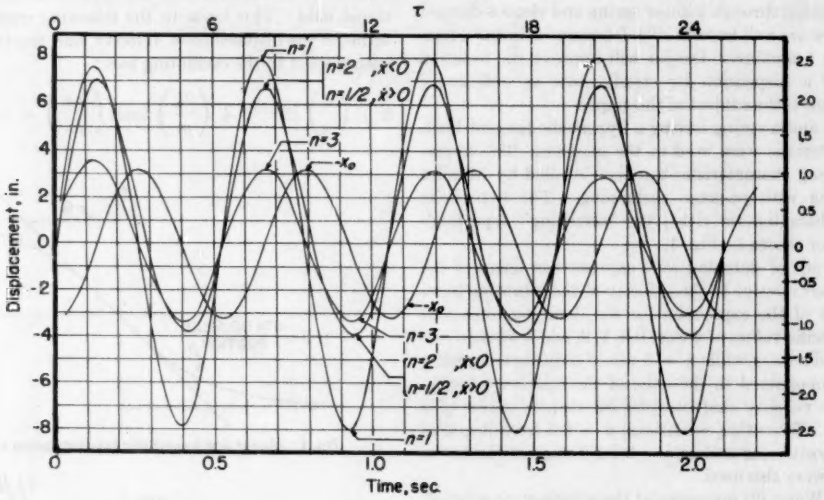


Fig. 3 Steady-state relative displacement-time curves for linear spring and various damper characteristics

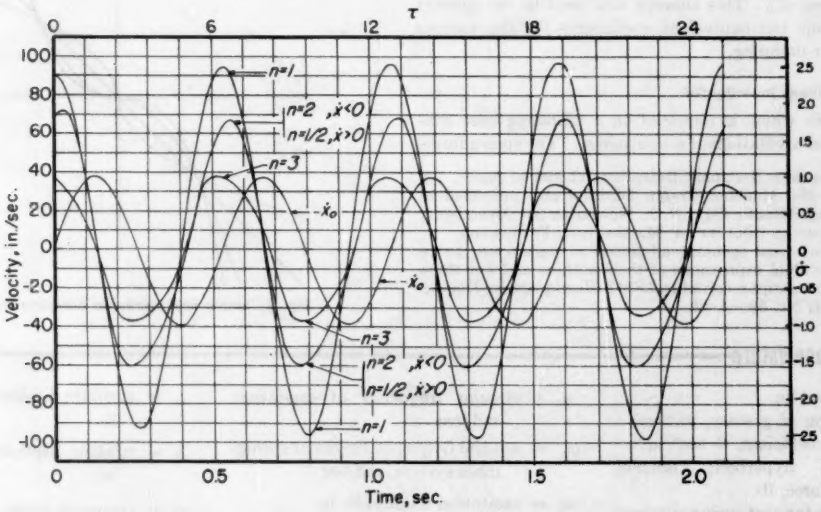


Fig. 4 Steady-state relative velocity-time curves for linear spring and various damper characteristics

and at the right for each figure. In terms of σ, the static mean position is -1.0125 for $r = 1.0$.

Discussion of Results

A General Comments

Only a small portion of the data collected appear in Fig. 3 through 10 because of publication space limitations. These curves present data corresponding to $\omega/\omega_n = 1$ for an excitation amplitude of $a_0 = 3.2$, or $\sigma_0 = 1$.

Relative displacement-time and relative velocity-time curves were used for the presentation of the steady-state data. The negative of the excitation displacement appears in Figs. 3 and 5. Since the absolute displacement is given by $x_1 = x - (-x_0)$, the difference between the relative displacement curves and the negative of the excitation displacement represents the absolute motion. The same procedure yields absolute velocities from Figs. 4 and 6.

The transient data were presented as absolute displacement-time and absolute velocity-time curves in Fig. 7 through 10. In this case at the end of one full cycle of the cosine wave, the excitation displacement was discontinued. Hence the curves beyond that time were absolute displacement and absolute velocity curves. The solutions for time less than one cycle were redrawn as absolute displacement and absolute velocity curves. Reversing the procedure of the foregoing paragraph, the relative displacement during the first cycle results from taking the difference between the positive excitation displacement and the absolute displacement curves.

Although only representative curves for $\omega/\omega_n = 1.0$ are presented, the results obtained for $\omega/\omega_n = 0.5$, 1.0, and 2.0 will be discussed. Before initiating this discussion, the criteria for one set of characteristics being superior to another set should be established.

For the condition being considered there are two possible means of comparison, one involves displacements and velocities, while the other involves forces. The force analysis in effect can be a comparison of the accelerations of the suspended body. In a

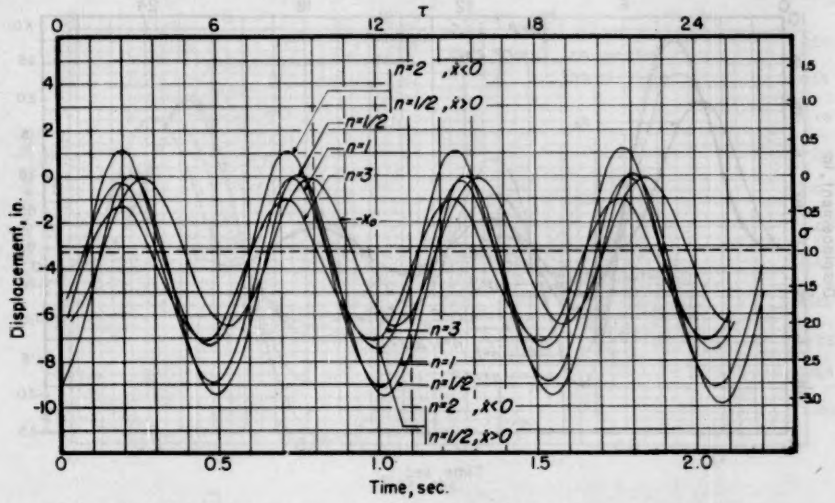


Fig. 5 Steady-state relative displacement-time curves for nonlinear spring and various damper characteristics

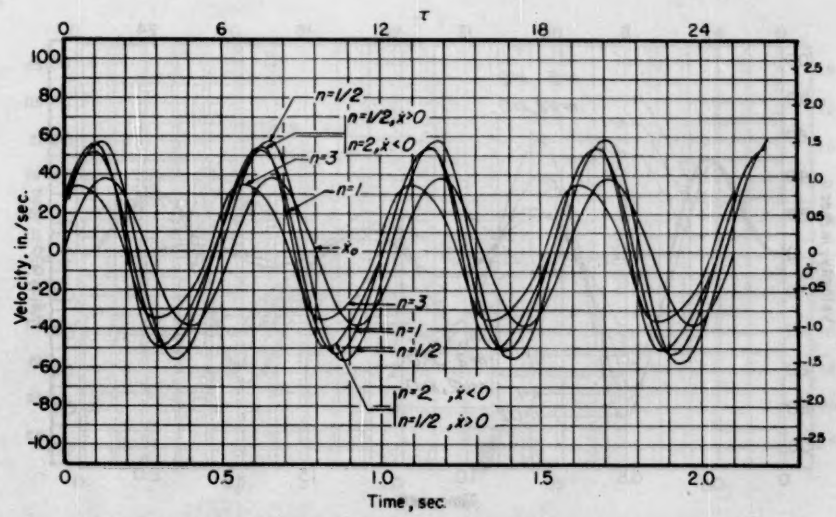


Fig. 6 Steady-state relative velocity-time curves for nonlinear spring and various damping characteristics

suspension system, such as a truck, tractor, or automobile seat, any or all of these quantities in addition to the rate of change of acceleration may be significant in establishing suitable vibration filter characteristics. Most of the analysis presented here is based on displacement and velocity considerations. Some attention is given to force considerations.

Because of the nature of the differential analyzer which was used, it was not possible to obtain accurate acceleration data nor was it possible to obtain accurate absolute displacement and absolute velocity data.

B Steady-State Vibration

1 Relative Displacement and Relative Velocity Data, $\omega/\omega_n = 0.5$. Using the linear spring the smallest amplitudes of relative displacement and relative velocity resulted when $n = 0.5$. For $n = 2(-), 0.5(+)$ they were practically the same as for viscous damping. The largest relative values resulted for $n = 3$. The various types of damping exhibited little difference in the time phase relation.

With the hyperbolic tangent spring characteristic the smallest amplitudes again resulted from $n = 0.5$ except for the largest excitation amplitude for which $n = 0.5(-), 2(+)$ gave the smallest value. The largest values were again obtained with $n = 3$. The major effect of using the nonlinear spring was the asymmetrical nature of the relative displacement and relative velocity-time diagrams. This effect was much more significant for $\omega/\omega_n = 0.5$ than at 1.0 or at 2.0.

Under these conditions the nonlinear spring resulted in relative displacement amplitudes approximately twice as large as those for the linear spring with the same damping characteristic.

2 Relative Displacement and Relative Velocity Data, $\omega/\omega_n = 1.0$. For the linear spring the double amplitudes of relative displacement were nearly identical for all types of nonlinear damping with $a_0 = 0.8$ which is the amplitude at which all dampers have

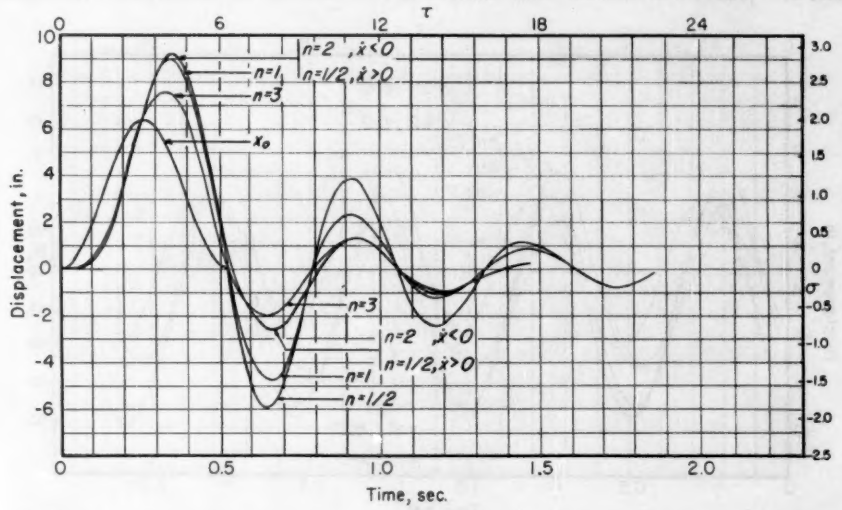


Fig. 7 Transient absolute displacement-time curves for linear spring and various damper characteristics

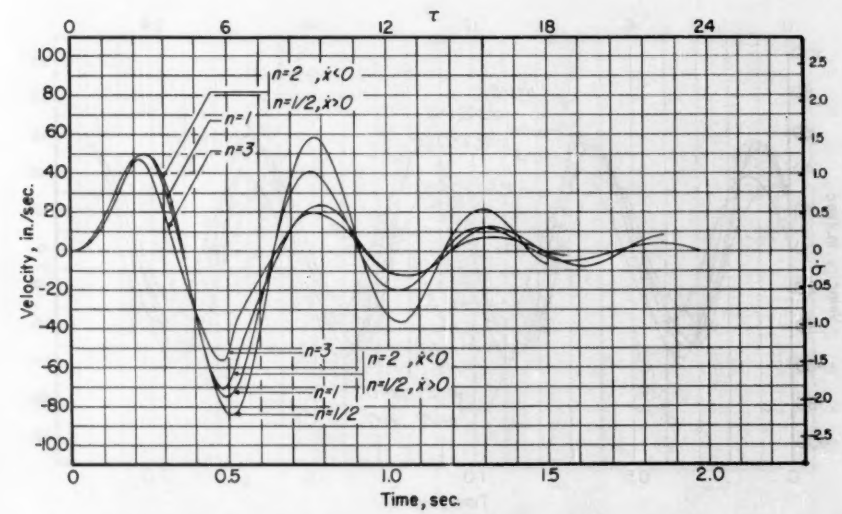


Fig. 8 Transient absolute velocity-time curves for linear spring and various damper characteristics

the same equivalent damping. This indicated that the concept proposed by Jacobsen [3] was reliable. Since the damping varies as some power of the relative motion amplitude and some power of the frequency for the various nonlinear dampers, there is only one frequency and one amplitude of motion for which the damping can be equivalent to that for the viscous damper.

With $n = 0.5$ the relative displacement amplitudes were small for small excitation amplitudes but became excessively large as a_0 was increased. This was because the damping was relatively large for small amplitudes and relatively small for large amplitudes. The reverse was true for $n = 2$ and $n = 3$ where the smallest a_0 yielded comparatively large relative displacements while the largest value of a_0 gave values only one half as large as for the viscous damper for $n = 2$ and two fifths as large for $n = 3$. When $n = 2(-), 0.5(+)$ the relative displacement amplitudes were smaller than for $n = 1$ but larger for $n = 2$ for a given value of a_0. Unsymmetrical damping of this type lead to an unsymmetrical relative displacement-time curve, see Fig. 3. The negative excursions of the mass relative to the base were less than the positive excursions for the two lowest values of a_0 but above this point the magnitudes reversed. Thus there was a shift in the mean position about which the body oscillated. For small a_0 the mean position shifted toward the base while for large a_0 it shifted away from the base.

A shift in the time at which peak relative displacement amplitudes occurred was also observed. For $n = 0.5$ the shift was to the left as a_0 was increased. For $n = 2$ and 3 the shift was to the right. For $n = 2(-), 0.5(+)$ the shift was self compensating and was essentially zero.

The relative velocity patterns were almost symmetrical amplitude wise but were unsymmetrical in shape for positive and negative values.

The nonlinear spring resulted in unsymmetrical relative displacement-time behavior for $n = 0.5$, 1, 2, and 3 but for $n = 2(-), 0.5(+)$ the pattern was almost symmetrical. This behavior was due to the compensating effect brought about by the

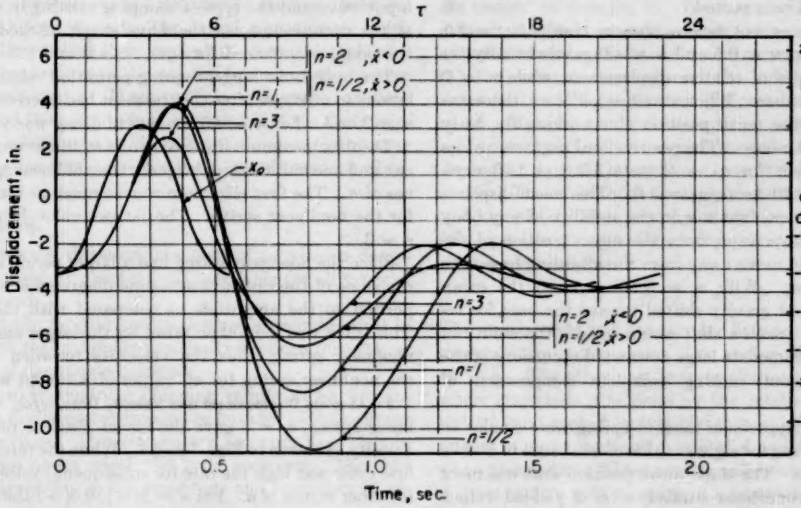


Fig. 9 Transient absolute displacement-time curves for nonlinear spring and various damper characteristics

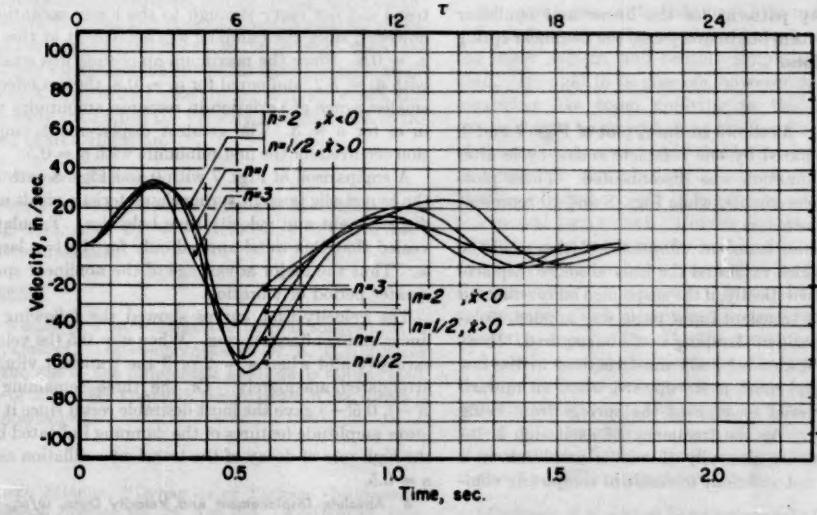


Fig. 10 Transient absolute velocity-time curves for nonlinear spring and various damper characteristics

nonlinear damping shifting the curves upward and the nonlinear spring shifting the curves downward.

Comparison of Figs. 3 and 5 shows the favorable decrease in relative displacement amplitudes which also were apparent at lower values of a_0. A comparison of Figs. 4 and 6, where the relative velocity curves appear, yields similar results. Part of the apparent effectiveness of the hyperbolic tangent spring in limiting the amplitude of relative displacement was due to the fact that the frequency ratio $\omega/\omega_n = 1.0$ was not the resonant frequency for the nonlinear spring when operating about the indicated static mean position. The frequency for maximum amplitude for this spring when used with viscous damping and with an excitation amplitude $a_0 = 2.0$ was found to be near $\omega/\omega_n = 0.7$. The maximum amplitude of x in this case was only 12 per cent higher than for $\omega/\omega_n = 1.0$ with $a_0 = 2.0$.

There is a favorable time shift of the wave shapes indicated in Fig. 6 as compared with those in Fig. 3, that is, a shift which makes the curves more nearly coincide with the $-x_0$ plot. The same behavior appears when comparing the relative velocities of Figs. 4 and 6 as would be expected.

3 Relative Displacement and Relative Velocity Data, $\omega/\omega_n = 2.0$. With the linear spring, $n = 0.5$ and $n = 1.0$ gave virtually the same double amplitudes of relative displacement while $n = 3$ resulted in the lowest values. When $n = 2(-), 0.5(+)$, there was a significant shift in the mean position about which the body oscillates relative to the base. The positive and negative values of relative displacement for $a_0 = 2$ were 3.5 and 1.33 compared with 4.71 and 3.30 for $\omega/\omega_n = 1.0$. This mean displacement would play an important role in the stability of any body mounted on such a suspension since the mean position of the center of gravity would move away from the vibrating base with increasing values of a_0. With $n = 0.5(-), 2(+)$ the mean position of the center of gravity shifted toward the base for increasing values of a_0. Such a shift would require the ability of the suspension to accommodate large downward excursions of the suspended body without causing complete compression of the springs.

The results for the hyperbolic tangent spring were similar to those for the linear spring for a given value of n in terms of double amplitude comparisons. The static mean position shift was more pronounced. Of all conditions studied, $n = 2$ yielded values most nearly equal to the amplitudes of the $-x_0$ curve.

In all cases the relative displacement amplitudes were less for the nonlinear spring.

The relative velocity patterns for the linear and nonlinear springs were similar in form but in every case the nonlinear spring resulted in smaller values.

C Transient Vibrations

1 General Comments. As shown in the x_0 plot of Figs. 7 and 9, the transient was generated by one complete cosine cycle after which the excitation function was disconnected. These plots represent absolute displacements, while Figs. 8 and 10 represent absolute velocities of the mass.

The criteria for superior transient vibration isolation would be indicated by motion which exhibited the least absolute displacement and velocity. Theoretically, if the suspended body remained at rest in space when a transient input pulse was applied, isolation would be complete and no damping would be required. However, for vehicles the suspended body must respond to the low frequency excitation and must move upward when an upward wave shape is encountered to prevent the springs from being compressed completely. As the frequency of excitation is increased, the excitations are generally of smaller amplitude in a practical case and are not sufficient to result in completely compressed springs.

Although a vehicle suspension has been used as an example here to illustrate a few points concerning transient vibration isolation, it is not intended to imply that these data can be applied to a system with two or more degrees of freedom.

Long continued oscillations following the transient pulse are usually undesirable, thus a damper is sought which will not contribute greatly to the absolute displacement during the transient pulse and which will attenuate the resulting oscillations following the transient pulse in a relatively short period of time.

2 Absolute Displacement and Velocity Data, $\omega/\omega_n = 0.5$. An analysis of tabulated values revealed that there were no large differences in the initial transient response for various types of damping when used with the linear spring. For $n = 0.5$ the free oscillations decayed rapidly but for $n = 2$ or 3 they decayed very slowly. Both types of unsymmetrical damping gave intermediate rates of decay. The two types of unsymmetrical damping were not considered for the steady-state condition since a reversal of sign merely inverted the wave shape. This was not the case with the transient vibration, however, since there was a single input wave and the type of damping existing in various portions of the cycle influenced the wave shape markedly. This effect was more pronounced for $\omega/\omega_n = 1.0$ and 2.0.

The hyperbolic tangent spring permitted relatively large oscillations to continue after the transient had been experienced when $n = 2$ or 3. For $n = 0.5$ the rate of decay was very high.

The first maximum displacement amplitude was approximately one and one half times as large for the nonlinear spring for all values of n. The first minimum was as much as four times as large for the nonlinear spring. The largest value being recorded for $n = 3$.

Since the nonlinear spring had a larger period of free vibration the slope of the displacement-time diagrams was smaller in proportion to the amplitude as compared with the linear spring. This effect was more than offset by the larger amplitudes for the nonlinear spring, thus the velocities recorded were larger for the nonlinear spring for all values of n except $n = 1.0$.

3 Absolute Displacement and Velocity Data, $\omega/\omega_n = 1.0$. For the linear spring, $n = 3$ gave the lowest absolute displacement and velocity as shown in Figs. 7 and 8. While the rate of decay for the first cycle was high the rate for subsequent cycles was lower than for other values of n. For $n = 2(-), 0.5(+)$ the free oscillations were nearly zero after two and one half cycles

For the nonlinear spring, $n = 3$ again yielded the lowest absolute displacement and velocity shown in Figs. 9 and 10. This trend did not carry through to the lower excitation amplitudes, however, since the damping was equivalent at this frequency for $a_0 = 0.8$. Since the maximum amplitude was smaller for $n = 3$ with $a_0 = 3.2$, and equal for $a_0 = 0.8$, the net effect was a much smaller range of variation in response amplitudes with variation in a_0 for $n = 3$. The greatest displacement amplitude variation occurred for the first minimum with $n = 0.5$.

A comparison of Figs. 7 with 9 and Figs. 8 with 10 shows that the hyperbolic tangent spring characteristic yields more favorable displacement and velocity-time behavior. Tabulated values revealed that this trend applied only for the two largest values of a_0. Thus the major advantage of the nonlinear spring lies in its greater period of vibration.

The velocity-time curves showed the following for both the linear and nonlinear spring. When $n = 0.5$ the velocities became excessive and when $n = 2$ or 3 the transient vibration was not attenuated adequately. Of the three remaining types, $n = 2(-), 0.5(+)$ gave the most desirable result since it combined the lower amplitude features of the damping indicated by $n = 2$ with the high rate of decay of the transient oscillation associated with $n = 0.5$.

4 Absolute Displacement and Velocity Data, $\omega/\omega_n = 2.0$. The

trends observed for $\omega/\omega_n = 1.0$ in general apply here for the linear spring. The rate of decay of the transient displacement-time curves for $n = 0.5$ is more rapid. The velocity-time curves are nearly twice as large for the first positive maximum.

D Force Transmitted to the Mass

A limited study of the force transmitted to the mass indicated that the nonlinear spring when used with nonlinear damping with $n = 3$ resulted in the least maximum force transmitted to the mass for $a_0 = 3.2$ with $\omega/\omega_n = 1.0$. For $a_0 = 2.0$, with $\omega/\omega_n = 1.0$ the maximum force transmitted to the mass was equal for $n = 3$ and $n = 2(-), 0.5(+)$. For $a_0 = 2.0$ with $\omega/\omega_n = 2.0$ the force transmitted was nearly equal for the cases given.

Conclusions

In evaluating the conclusions drawn from this investigation it is necessary to remember that a limited frequency range has been investigated. Further, the various combinations of springs and dampers have been studied for one coefficient of damping only. The particular coefficient selected may not be the best choice for the range of conditions studied. The damping coefficients which were used were those which gave equivalent viscous damping for a frequency corresponding to the natural frequency of a system containing a linear spring. These coefficients depended on both the frequency and the amplitude of the motion of the body relative to the motion of the base. Recall that the best compromise is sought for isolation from both steady-state and transient vibrations.

Steady-State Conditions

1 For steady-state vibration, the equivalent viscous damping concept proposed by Jacobsen [3] is reliable for all types of damping considered.

2 The hyperbolic tangent spring displays a desirable shift in the time at which the maximum amplitude develops for $\omega/\omega_n = 1.0$ for all types of damping considered.

3 Nonlinear unsymmetrical damping results in an unsymmetrical relative displacement-time wave form for both the linear and nonlinear springs. This asymmetry amounts to a shift in the mean position about which the body oscillates. The shift varies in sense according to the amplitude and frequency of excitation.

Transient Conditions

4 The nonlinear unsymmetrically loaded spring increases the period of the free oscillations and thus reduces the relative velocities for excitation frequencies equal to or greater than the resonant frequency for the equivalent linear spring-mass system. Below resonance the increased velocity amplitude, which this spring permits, more than offset the velocity reduction due to the increased period of free vibration.

Steady-State and Transient Conditions

5 For all combinations of steady-state and transient vibrations, damping designated by $n = 3$ or $n = 2(-), 0.5(+)$ offered the best compromise.

Acknowledgments

The author wishes to acknowledge the helpful suggestions of Dr. Nils O. Myklestad. The co-operation of Dr. A. T. Nordsieck in permitting the use of the differential analyzer is appreciated. The assistance of James R. Young in the analysis of the forces is gratefully acknowledged.

References

1 Raymond D. Mindlin, "Dynamics of Package Cushioning," *Bell System Technical Journal*, vol. 24, Nos. 3–4, July–October, 1945, pp. 353–461.

2 R. M. Rosenberg and A. J. Wang, "Periodic Solutions of a Nonlinear Differential Equation," Univ. of Wash. Eng. Exp. Sta. Bulletin No. 118, Aero. Series No. 1, Univ. of Washington Press, Seattle, Wash., December, 1950, pp. 77–87.

3 Lydik S. Jacobsen, "Steady Forced Vibation as Influenced by Damping," TRANS. ASME, vol. 52, Part I, 1930, APM-52-15, pp. 169–181.

4 Arnold T. Nordsieck, "A Synchro-Operated Differential Analyzer," published in *Digital and Analog Computers and Computing Methods*, Symposium at the 18th Applied Mechanics Division Conference of the ASME held at the University of Minnesota, June 18–20, 1953.

5 Will J. Worley, "An Analytical Investigation of Mechanical Vibration Filters With Nonlinear Characteristics," PhD thesis, Univ. of Illinois, Urbana, Ill., 1952.

DISCUSSION

W. J. Cunningham[2]

The problem considered in this paper is similar to many other nonlinear problems in that, while the system being studied appears superficially to be simple, a considerable number of parameters are needed to describe it. Because of this large number, it is difficult to obtain a simple presentation of the operation of the system and of the effects of the various parameters. In Equation (1), for example, there appears the independent variable t, the dependent variable x, and the eight parameters $m, c, n, P_0, k_0, g, a_0$, and ω. By normalizing the quantities, as is done in Equation (2), a simplification is obtained with the independent variable τ, the dependent variable σ, and now only five parameters, r, q_a, n, u, and h. The effects of these five parameters need to be known.

A completely linear system would be described by Equation (2) if parameter n becomes unity and P_0 becomes infinite. The resulting equation is

$$r^2\ddot{\sigma} + q\dot{\sigma} + \sigma = -h + \cos\tau$$

The steady-state solution for this linear equation consists of the superposition of two parts. One part is the steady deflection, $\sigma_s = -h$. The other part is an oscillation, $\sigma_0 = A\cos(\tau - \phi)$, where amplitude A is given by the relation $A = [q^2 + (1 - r^2)^2]^{-1/2}$ and angle ϕ is given by $\tan\phi = q/(1 - r^2)$. It is a simple matter to plot a family of curves of A and ϕ as functions of r with q as a parameter. These are the well-known resonance curves for a linear system with damping as a parameter. These curves present in an easily visualized manner all that is significant about the operation of the linear system. What is needed is some analogous procedure for the nonlinear system.

Representation of the behavior of the nonlinear system is made difficult because of the extra parameters needed beyond those of the linear system, and because simple superposition cannot be used. It ought to be possible, however, to plot resonance curves somewhat like those applying to the linear system. These curves should be skewed along the frequency (i.e., r) axis, because of the nonlinear spring, and should show the possibility of jumps in amplitude characteristic of the Duffing equation. Several families of curves would be needed to account for the effects of the several parameters. Curves of this kind would present information in a more concise and readily understandable manner than do the curves in the paper showing instantaneous values of the deflections and velocities. It would be worth while to devote considerable effort to arrange a concise presentation of the large amount of data produced by the computing machine.

As a minor point, it appears that the type of spring described by the hyperbolic tangent function of the paper is rather uncommon. Mindlin (reference [1] of the paper) observes that a strut subjected to a longitudinal force does, indeed, approximate this characteristic. Most common springs, however, are likely to be of the hard rather than the soft type.

[2] Professor of Electrical Engineering, Yale University, New Haven, Conn.

Author's Closure

The first two paragraphs of the discussion state only that Eqs. (1) and (2) are sufficiently versatile to include the linear condition. Indeed, the linear conditions appear in Figs. 1 and 2 of the paper. The linear damper condition, $n = 1$, was used to check the validity of the analog computer solution as well as to obtain a reference index in Figs. 3, 4, 7, and 8 for the linear spring and in Figs. 5, 6, 9, and 10 for the nonlinear spring.

The statements of the third paragraph anticipate the contents of an Air Force contract under the direction of the writer which was initiated in June, 1959, and for which a final report will appear in the summer of 1960.

BERNARD W. SHAFFER
Professor of Mechanical Engineering, Mechanical Engineering Department, New York University, New York, N. Y. Mem. ASME

IRVIN KRAUSE
New York University, New York, N. Y. Lecturer (on leave of absence), Department of Drafting, City College of New York. Assoc. Mem. ASME

Refinement of Finite Difference Calculations in Kinematic Analysis

First and second finite-difference expressions are derived for the first three derivatives of displacement with respect to time. The numerical results obtained by applying these difference expressions to a specific cam-follower mechanism are found to be in good agreement with the values obtained by successive differentiation of the displacement-time equation. Procedures for evaluating and improving the numerical results are also discussed.

Introduction

THE METHOD OF FINITE DIFFERENCES, which has long been used to find numerical solutions to differential equations, has only very recently been applied to problems in kinematics. The advantages that this method has over analytical and graphical methods were discussed in some of the previous papers [1, 2, 3].[1] However, it also has the disadvantage of providing numerical results which deviate somewhat from exact values. Most kinematicians realize that the deviation may be diminished by reducing the space or time interval used in a specific analysis. Few realize, however, that, when the afore-mentioned method of error reduction is not practical, other more advanced methods may be used. Some of these methods will be discussed in the present paper.

It has also been observed that kinematicians have restricted their analyses of mechanisms by the use of finite difference expressions solely to the first and second derivatives with respect to time. The third derivative, which is of particular importance in cam design, may also be evaluated in a similar manner. It, too, will be discussed in the present paper.

Finally, the finite difference expressions will be applied to the analysis of a cam and follower mechanism in order to demonstrate the use of the procedures described in the paper. Since the method is very general, it may also be applied to the analysis of other more complicated mechanisms.

The mathematical technique to be discussed is not original. Nevertheless, it is worth while to discuss the technique because its application to problems in kinematics has not been fully appreciated.

The Finite-Difference Expressions

The Cartesian co-ordinates (x, y) of any point in a mechanism may be expressed in terms of the angular co-ordinate θ of the driving link, or directly in terms of the time co-ordinate t, for a prescribed angular velocity of the driving link. The functions $x(t)$ and $y(t)$ are continuous and so are its derivatives. One may, therefore, write Taylor's series for y as a function of time $y(t)$[2] either in the form

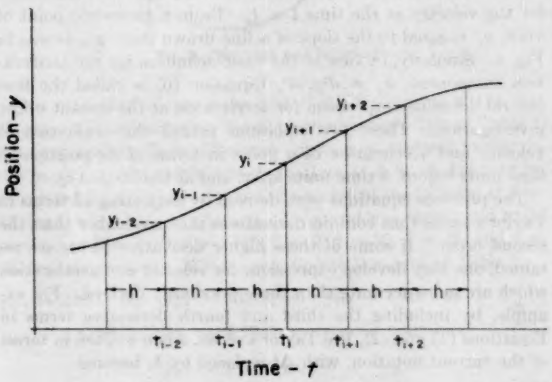


Fig. 1 A displacement-time curve

$$y(t + \Delta t) = y(t) + (\Delta t)\frac{dy(t)}{dt} + \frac{(\Delta t)^2}{2!}\frac{d^2y(t)}{dt^2} + \frac{(\Delta t)^3}{3!}\frac{d^3y(t)}{dt^3} + \dots \quad (1)$$

or in the form

$$y(t - \Delta t) = y(t) - (\Delta t)\frac{dy(t)}{dt} + \frac{(\Delta t)^2}{2!}\frac{d^2y(t)}{dt^2} - \frac{(\Delta t)^3}{3!}\frac{d^3y(t)}{dt^3} + \dots \quad (2)$$

for the purposes of comparing the values of the function at t and $(t + \Delta t)$, or t and $(t - \Delta t)$, respectively.

For a given mechanism, the function $y(t)$ is known either algebraically or numerically. In either case, numerical values of the function may be determined at uniform time intervals, h time units apart, as indicated in Fig. 1.

It is convenient, in the analysis to follow, to study the displacement-time curve one point at a time. The procedure is facilitated by denoting the particular time to be investigated by t_i and by denoting the corresponding space co-ordinate by y_i. Similarly, the time "stations" to the left of t_i are labeled t_{i-1}, t_{i-2}, . . . and those to the right of t_i are labeled t_{i+1}, t_{i+2}, . . . The corresponding space co-ordinates are denoted by y_{i-1}, y_{i-2}, . . . and y_{i+1}, y_{i+2}, . . . respectively.

With this notation, and with the time interval Δt set equal to h, Equation (1) may be written

$$y_{i+1} = y_i + hy_i' + \frac{h^2}{2!}y_i'', \quad (3)$$

[1] Numbers in brackets designate References at end of paper.

[2] Analogous expressions may be written for $x(t)$. The resulting finite difference expressions are identical except that x and y are interchanged.

Contributed by the Machine Design Division and presented at the Annual Meeting, Atlantic City, N. J., November 29–December 4, 1959, of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS.

NOTE: Statements and opinions advanced in papers are to be understood as individual expressions of their authors and not those of the Society. Manuscript received at ASME Headquarters, July 6, 1959. Paper No. 59—A-53.

while Equation (2) may be written

$$y_{i-1} = y_i - hy_i' + \frac{h^2}{2!} y_i'', \quad (4)$$

where the higher-order terms have been neglected and the primes indicate differentiation with respect to time. Subtraction of Equation (4) from Equation (3) leads to

$$y_i' = \frac{y_{i+1} - y_{i-1}}{2h}, \quad (5)$$

whereas their addition leads to

$$y_i'' = \frac{y_{i-1} - 2y_i + y_{i+1}}{h^2}. \quad (6)$$

In view of the basic definition for the velocity component $v_y = dy/dt$, Equation (5) is called the first central difference expression for the velocity at the time $t = t_i$. From a geometric point of view, y_i' is equal to the slope of a line drawn from y_{i-1} to y_{i+1} in Fig. 1. Similarly, in view of the basic definition for the acceleration component, $a_y = d^2y/dt^2$, Equation (6) is called the first central difference expression for acceleraton at the instant under investigation. These two equations permit the evaluation of velocity and acceleration of a point in terms of its positions h time units before, h time units after, and at the instant t_i.

The previous equations were derived by neglecting all terms in Taylor's series that contain derivatives that are higher than the second order. If some of these higher derivative terms are retained, one may develop expressions for velocity and acceleration which are more accurate than those previously derived. For example, by including the third and fourth derivative terms in Equations (1) and (2), the Taylor's series, when written in terms of the current notation, with Δt replaced by h, become

$$y_{i+1} = y_i + hy_i' + \frac{h^2}{2!} y_i'' + \frac{h^3}{3!} y_i''' + \frac{h^4}{4!} y_i^{iv} \quad (7)$$

and

$$y_{i-1} = y_i - hy_i' + \frac{h^2}{2!} y_i'' - \frac{h^3}{3!} y_i''' + \frac{h^4}{4!} y_i^{iv}. \quad (8)$$

From these, it is to be noted that

$$y_{i+1} - y_{i-1} = 2hy_i' + \frac{2h^3}{3!} y_i'''. \quad (9)$$

If the original Taylor's series are rewritten in terms of the current notation, but with Δt now replaced by $2h$, Equations (1) and (2) become

$$y_{i+2} = y_i + (2h)y_i' + \frac{(2h)^2}{2!} y_i'' + \frac{(2h)^3}{3!} y_i''' + \frac{(2h)^4}{4!} y_i^{iv} \quad (10)$$

and

$$y_{i-2} = y_i - (2h)y_i' + \frac{(2h)^2}{2!} y_i'' - \frac{(2h)^3}{3!} y_i''' + \frac{(2h)^4}{4!} y_i^{iv} \quad (11)$$

where, in keeping with the current notation, $y(t + 2h) = y_{i+2}$ and $y(t - 2h) = y_{i-2}$. The difference between Equations (10) and (11) is

$$y_{i+2} - y_{i-2} = 4hy_i' + 2\frac{(2h)^3}{3!} y_i'''. \quad (12)$$

From Equations (9) and (12), it may be established that

$$y_i' = \frac{y_{i-2} - 8y_{i-1} + 8y_{i+1} - y_{i+2}}{12h}. \quad (13)$$

Also, from the sum of Equations (7) and (8), and the sum of Equations (10) and (11), it is found that

$$y_i'' = \frac{-y_{i-2} + 16y_{i-1} - 30y_i + 16y_{i+1} - y_{i+2}}{12h^2}. \quad (14)$$

In view of the physical meaning of dy/dt and d^2y/dt^2, the last two equations describe the velocity and acceleration of a point in a mechanism. They were obtained by including terms in Taylor's series that are of higher order than those retained in the earlier derivations. Thus Equation (13) is called the second central difference expression for velocity, and Equation (14) is called the second central difference expression for acceleration at the instant $t = t_i$. These finite difference expressions require information concerning the position of the point up to $2h$ time units from the instant under consideration.

It is also possible to derive a finite difference expression for the third derivative of displacement with respect to time. If Equations (9) and (12) are solved simultaneously for y_i''', it is found that

$$y_i''' = \frac{-y_{i-2} + 2y_{i-1} - 2y_{i+1} + y_{i+2}}{2h^3}. \quad (15)$$

Equation (15) is the first central difference approximation for d^3y/dt^3, which is sometimes called the pulse or jerk.

If the fifth and sixth derivative terms are included in Equations (1), (2), (7), (8), (10), and (11), and Equations (1) and (2) are rewritten in terms of the current notation with Δt replaced by $3h$, then the second difference approximation for the third derivative may be derived. When these six equations are solved for y_i''', one finds that

$$y_i''' = \frac{y_{i-3} - 8y_{i-2} + 13y_{i-1} - 13y_{i+1} + 8y_{i+2} - y_{i+3}}{8h^3}. \quad (16)$$

If the same six equations were used to solve for y_i' and y_i'', the resulting expressions would be the third central difference approximations of velocity and acceleration.

Finite difference expressions may be derived more concisely if operator notation is used for the derivatives and the difference expressions [4, 5]. Use of operator notation also reduces greatly the tedious work inherent in the progressive improvement of numerical approximations. However, because this paper is directed toward design engineers who are more interested in underlying principles than mathematical neatness, it is the authors' feeling that the use of operation notation would tend to complicate rather than simplify.

Application of Finite Difference Expressions

In the present section, a specific problem is analyzed kinematically by using the first and second central difference expressions. The results are then compared with those obtained from an exact solution. This approach is used to demonstrate the application of the finite difference expressions and to observe the accuracy that is obtained.

Consider the circular plate cam and follower shown in Fig. 2. The cam, of radius r, is rotated about the axis O which is a distance e from the geometric center C.

The y-axis is chosen so that it originates at the center of rotation of the cam and is parallel to the direction of motion of the follower. The angular position of the cam, represented by θ, is measured clockwise from the y-axis to the line OC. From geometric considerations, one may describe the position of the tip of the follower in terms of the physical dimensions of the cam and its angular position θ as

$$y = e\cos\theta + r\sqrt{1 - \left(\frac{e}{r}\sin\theta\right)^2}. \quad (17)$$

The velocity of the follower is the first derivative of Equation (17) with respect to time. Since the angular velocity, $\omega = d\theta/dt$,

$$v_y = \frac{dy}{dt} = -e\omega\left[\sin\theta + \frac{e\sin 2\theta}{2r\sqrt{1 - \left(\frac{e}{r}\sin\theta\right)^2}}\right]. \quad (18)$$

The acceleration of the follower is the second derivative of Equation (17) with respect to time. If ω is specified as constant, $d\omega/dt = 0$ and

$$a_y = \frac{d^2y}{dt^2} =$$

$$-e\omega^2\left\{\cos\theta + \frac{\left[\frac{e}{r} - \left(\frac{e}{r}\right)^3\right]\cos 2\theta + \left(\frac{e}{r}\right)^3\cos^4\theta}{\left[1 - \left(\frac{e}{r}\sin\theta\right)^2\right]^{3/2}}\right\}. \quad (19)$$

The third derivative of the follower displacement with respect to time is, in view of Equation (17),

$$\frac{d^3y}{dt^3} = e\omega^3\left\{\sin\theta + \frac{2A\sin 2\theta}{\left[1 - \left(\frac{e}{r}\sin\theta\right)^2\right]^3}\right\}, \quad (20)$$

where

$$A = \left[1 - \left(\frac{e}{r}\sin\theta\right)^2\right]^{3/2}\left\{\left[\frac{e}{r} - \left(\frac{e}{r}\right)^3\right] + \left(\frac{e}{r}\right)^3\cos^2\theta\right\} - \left\{\left[\frac{e}{r} - \left(\frac{e}{r}\right)^3\right]\cos 2\theta + \left(\frac{e}{r}\right)^3\cos^4\theta\right\}\left\{3\left(\frac{e}{2r}\right)^2\left[1 - \left(\frac{e}{r}\sin\theta\right)^2\right]^{1/2}\right\}$$

For purposes of a specific analysis, the cam shown in Fig. 2 is assumed to have an eccentricity e equal to 0.6 in., a radius r equal to 1.6 in., and an angular velocity ω equal to 100 rad/sec. The displacement of the follower and its first three derivatives with respect to time may be found by substituting the appropriate dimensions into Equations (17), (18), (19), and (20), respectively.

This was done for every 15 deg of rotation and the results appear in Table 1. The results are tabulated only for the first 180 deg of rotation because of the symmetry of the problem.

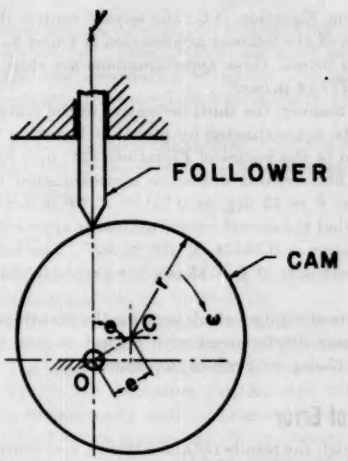


Fig. 2 The cam and follower

One may also calculate the first three derivatives of the displacement by substituting the displacement-time date of Table 1 into appropriate finite-difference expressions derived earlier in the paper. For example, the derivatives at $\theta = 15$ deg may be calculated by designating y_i as the position of the follower at $\theta = 15$ deg; while y_{i-3}, y_{i-2}, y_{i-1}, y_{i+1}, y_{i+2}, and y_{i+3} correspond to its position at $\theta = -30$, $\theta = -15$, $\theta = 0$, $\theta = 30$, $\theta = 45$, and $\theta = 60$ deg, respectively. Thus, according to the data of Table 1,

$$\left.\begin{array}{lll} y_{i-3} = 2.0912 \text{ in.} & y_{i-1} = 2.2000 \text{ in.} & y_{i+2} = 1.9670 \text{ in.} \\ y_{i-2} = 2.1720 \text{ in.} & y_i = 2.1720 \text{ in.} & y_{i+3} = 1.8133 \text{ in.} \\ & y_{i+1} = 2.0912 \text{ in.} & \end{array}\right\} \quad (21)$$

When the station i corresponds to $\theta = 15$ deg notice that the numerical values of y_{i-2} and y_i are equal because of the symmetry of the problem.

According to the first central difference approximation, Equation (5), the velocity of the follower is -20.779 in/sec, when the displacements are prescribed by Equation (21) and $h = 0.00261799$ sec. According to Equation (13), which is the second central difference approximation, the follower velocity at $\theta = 15$ deg is -21.180 in/sec. These values have also been tabulated in Table 1 and compare favorably with the exact value of -21.181 in/sec.

Similarly, the acceleration of the follower may also be found by using the finite difference expressions derived earlier. When the numerical values of Equations (21) are introduced into Equation (6), the first central difference approximation for acceleration of the follower at $\theta = 15$ deg is found to be -7703.7 in/sec^2;

Table 1

Pos. of cam, θ, degrees	Pos. of follower, y, inches	dy/dt 1st cent. difference, in/sec	dy/dt 2nd cent. difference, in/sec	dy/dt Exact method, in/sec	d^2y/dt^2 1st cent. difference, in/sec^2	d^2y/dt^2 2nd cent. difference, in/sec^2	d^2y/dt^2 Exact method, in/sec^2	$d^3y/dt^3 \times 10^{-6}$ 1st cent. difference, in/sec^3	$d^3y/dt^3 \times 10^{-6}$ 2nd cent. difference, in/sec^3	$d^3y/dt^3 \times 10^{-6}$ Exact method, in/sec^3
0	2.2000	0.000	0.000	0.000	−8170.5	−8248.4	−8250.0	0.00000	0.00000	0.00000
15	2.1720	−20.779	−21.180	−21.181	−7703.7	−7779.0	−7773.4	0.35110	0.36434	0.36031
30	2.0912	−39.152	−39.893	−39.919	−6332.2	−6386.9	−6404.0	0.64926	0.67434	0.67388
45	1.9670	−53.075	−54.043	−54.094	−4304.1	−4335.7	−4330.9	0.84711	0.87985	0.88897
60	1.8133	−61.192	−62.236	−62.263	−1896.7	−1894.3	−1880.7	0.91398	0.95508	0.95452
75	1.6466	−63.044	−63.977	−63.990	48.2	52.2	51.3	0.81645	0.84989	0.84603
90	1.4832	−59.301	−59.970	−60.000	2378.2	2439.0	2427.2	0.58517	0.59214	0.60000
105	1.3361	−51.547	−51.920	−51.921	3545.4	3597.7	3619.0	0.32602	0.31488	0.31309
120	1.2133	−41.559	−41.686	−41.660	4085.3	4126.6	4119.3	0.11146	0.08987	0.08741
135	1.1185	−30.806	−30.787	−30.759	4129.0	4143.6	4154.5	−0.01672	−0.03831	−0.04043
150	1.0520	−20.168	−20.101	−20.081	3997.7	4001.4	3938.3	−0.05852	−0.07245	−0.07388
165	1.0129	−9.931	−9.880	−9.879	3822.6	3812.9	3817.7	−0.04458	−0.05225	−0.04972
180	1.0000	0.000	0.000	0.000	3764.2	3754.6	3750.0	0.00000	0.00000	0.00000

whereas from Equation (14), the second central difference approximation of the follower acceleration is found to be -7779.0 in/sec^2. As before, these approximations are close to the exact value of -7773.4 in/sec^2.

In a like manner, the third derivative of the follower displacement may be approximated by means of Equation (15) or (16). Substitution of the values of Equations (21) into Equation (15) yields the first central difference approximation of the third derivative at $\theta = 15$ deg, as 0.35110×10^6 in/sec^3. Equation (16) shows that the second central difference approximation of the third derivative is 0.36434×10^6 in/sec^3. The exact value of the third derivative at $\theta = 15$ deg is equal to 0.36031×10^6 in/sec^3.

The results of similar calculations for the first three derivatives of the follower displacement with respect to time are listed in Table 1 for 15-deg intervals of cam rotation.

Reduction of Error

Even though the results obtained from a first central difference calculation are reasonably close to exact numerical values, the errors can be reduced. One method is to reduce the time interval initially used. A second method, demonstrated in the previous sections, calls for the inclusion of higher-order terms of the Taylor's series. A third method is to extrapolate previously obtained results.

A logical method of extrapolation may be obtained [6] by reverting to the original equation upon which the finite difference approximations are based. Suppose $y'_{i,h}$ and $y'_{i,h/2}$ represent numerical results obtained for the first central-difference approximations of velocity, at $t = t_i$ when Δt is equal to h and $h/2$, respectively. The first central difference approximations for velocity, you will recall, were derived by neglecting all those terms in Equations (3) and (4) that contained derivatives beyond the second. If the derivation were now repeated retaining the third derivative, Equation (5) would read

$$y_i' = y'_{i,h} - \frac{2h^2}{3!} y_i''', \tag{22}$$

for $\Delta t = h$, and

$$y_i' = y'_{i,h/2} - \frac{2\left[\frac{h}{2}\right]^2}{3!} y_i''', \tag{23}$$

for $\Delta t = h/2$. Since $y'_{i,h}$ and $y'_{i,h/2}$ are quantities available from previous calculations, an even closer approximation of the exact velocity, y_i' may be computed by the simultaneous solution of Equations (22) and (23).

Similarly, once *three* first central difference approximations have been calculated at $t = t_i$ by using three different successively smaller time intervals, these approximations can also be easily extrapolated by substituting into a newly derived set of *three* equations analogous to Equations (22) and (23). The new set is derived by including the third and fifth derivative terms in Equation (5) and by rewriting it in terms of *each of the three* successively smaller time intervals.

The same technique may be used to improve the calculations for the second and third derivatives. It must be noted, however, that it is possible to use the extrapolation procedure only if values of displacement are available at each of the time stations which are now required to be closer together than they were for the first central difference calculation.

Concluding Remarks

It has been shown that the finite-difference expressions developed in the previous sections may be used to evaluate numerically the first three derivatives of the follower displacement with respect to time. Two expressions, representing a first and second approximation, were presented in each case.

When the finite difference expressions are applied to a specific cam mechanism and the results are compared with those obtained from the exact solution, it is found that the first central difference expressions lead to numerical results that are reasonable for engineering studies, whereas the second central difference expressions lead to even better agreement. In a practical case, where the exact solution is not known in advance, closeness of the results obtained by using the first and second-order finite difference expressions indicates relative accuracy of the numerical results and convergence to the exact solution. In cases where the numerical results are not sufficiently accurate, they may be improved by including still higher-order terms in the Taylor's series expansion by using smaller time intervals, or by extrapolating previous calculations.

It is significant to observe from the data listed in Table 1 that the results found by using finite difference expressions are in close agreement with results found by using analytical expressions, obtained by successive differentiation of the displacement-time relation, even though crank angle intervals as large as 15 deg were used. Obviously, one need not resort to infinitesimal increments in order to obtain reasonable numerical results.

Even though the present paper is directed toward a discussion of the refinement of numerical calculations, the authors feel it is worth while at this time to briefly mention some of the additional finite difference concepts which may also be used in kinematic analysis. These concepts are described in texts dealing with the finite difference method [4, 5]. One is concerned with finite difference expressions written entirely in terms of stations in advance of the one under investigation. They are known as the forward finite difference expressions. Another is concerned with finite difference expressions written entirely in terms of stations behind the one under investigation. They are known as backward finite difference expressions. Also, the restriction of equally spaced stations may be removed so that finite difference expressions can be written in terms of unequally spaced stations. Finally, operator notation may be used, which greatly increases the efficiency of computation of successively better approximations.

It is apparent that the finite difference technique, currently being used in other engineering fields, can also be fruitfully applied to the analysis of mechanisms.

Acknowledgment

The authors wish to thank the National Science Foundation for making this paper possible by its support through Science Faculty Fellowship 69178.

References

1 B. W. Shaffer and I. Krause, "Kinematic Analysis of a Specific Point in a Mechanism by the Use of Finite Difference Expressions," N.Y.U., Research Report, November, 1957.

2 Ray C. Johnson, "Development of a High Speed Indexing Mechanism," *Machine Design*, Sept. 4, 1958.

3 Ray C. Johnson, "The Dynamic Analysis and Design of Relatively Flexible Cam Mechanisms Having More Than One Degree of Freedom," TRANS. ASME, Series B, JOURNAL OF ENGINEERING FOR INDUSTRY, vol. 81, 1959, p. 323.

4 M. G. Salvadori and M. Baron, "Numerical Methods in Engineering," Prentice-Hall, Inc., New York, N. Y., 1952.

5 F. B. Hildebrand, "Introduction to Numerical Analysis," McGraw-Hill Book Company, Inc., New York, N. Y., 1956.

6 L. F. Richardson, "The Approximate Arithmetical Solution by Finite Differences of Physical Problems Involving Differential Equations With Application to the Stresses in a Masonry Dam," *Philosophical Trans., Royal Society, London*, vol. 210, 1911.

DISCUSSION

Thomas P. Goodman[3]

The authors are to be commended for their lucid engineering presentation of an important mathematical technique. In view of the application of this technique to cam analysis and design, it would be of interest to have the authors' comments on how it applies:

(*a*) At transition points on a cam profile such as the point where a rise portion joins a dwell portion.

(*b*) For cams having random machining errors such that the error-reduction process described in the paper does not converge except for extremely small values of h.

It should be noted that the finite-difference equations can be applied not only to scalars but also to vector quantities, simply by interpreting y and its derivatives as vectors in Equations (1) through (16) of the paper. (The additions and subtractions in these equations then become vector additions and subtractions.) J. A Hrones[4,5] suggested the use of first finite-difference equations for vector velocities and accelerations, and the accuracy of his method could undoubtedly be improved by using the higher finite-difference equations given in the paper. The use of finite-difference equations would be especially advantageous in finding the derivative of the acceleration vector, for which vector-polygon solutions become rather complicated.[6]

Harold A. Rothbart[7]

The authors are to be congratulated on an excellent contribution to the study of machinery and mechanisms. Particular interest is noted in the comparison of accuracy of the finite difference method. The subject of high-speed operation of machinery has special need for the use of the method shown in this paper, since the value of the second derivative (acceleration y'') of compound mechanism cannot conveniently be found analytically. The equations are often too cumbersome to differentiate, and conventional graphical solution of velocities y' and acceleration y'' are quite time-consuming. The finite difference method yields all values easily and directly. It is interesting to note that the third derivative (y''' pulse or jerk) has much less practical significance than the acceleration y'' curve. Its predominant value is in appraising the second derivative curve for dynamic consideration.

The writer would appreciate the authors' thoughts on the relationship between the accuracy of the finite difference and input cam or crank position.

Authors' Closure

The authors appreciate the illuminating discussions presented by Dr. T. Goodman and Dr. H. Rothbart.

It was stated in the paper that, before a function can be expanded in a Taylor series, the function must possess continuous derivatives. The implication is that all derivatives exist and are continuous. Yet, in the derivation that followed only some of the derivatives were actually used. For example, in evaluating the first central difference expressions, terms beyond the fourth derivative were omitted as being of higher order; and in evaluating the second central difference expressions, terms beyond the sixth derivative were omitted for similar reasons. Thus no difficulty would be expected in applying the central difference expressions to a function in which not all the higher ordered terms are continuous.

A function in the neighborhood of a transition point of a cam profile is one which may not have all derivatives. Thus difficulty should be expected in applying the central difference expressions when the nonexisting derivatives are of low enough order to have been used in the derivations. In the afore-mentioned case, the difficulty can be avoided by using forward or backward difference expressions [4, 5], since all the required derivatives of the function may exist to either side of the transition point. On the other hand, no difficulty will occur when only the higher ordered terms do not exist, in which case the central difference expressions of the present paper may, of course, be used.

The finite difference method was proposed as a means of analyzing basic cam motion just as one may analyze basic motion by suitable differentiation of a displacement-time function. In neither case does one consider machining errors, an effect which is important but not considered in an initial investigation.

It may be that one can incorporate the tolerances associated with machining errors to determine upper and lower bounds to the first three derivatives of displacement with respect to time by using a very small time interval h together with the ideal displacement-time relations. However, the problem is sufficiently complex so that an exact statement of procedure cannot be made without additional study.

---

[3] Kinematics Engineer, General Engineering Laboratory, General Electric Company, Schenectady, N. Y. Assoc. Mem. ASME.

[4] J. A. Hrones, "A Job-Shop Approach to Mechanism Analysis," *Machine Design*, vol. 26, 1954, pp. 183–190.

[5] J. A. Hrones and G. L. Nelson, "Analysis of the Four-Bar Linkage," The Technology Press and John Wiley & Sons, Inc., New York, N. Y., 1951 (page xix).

[6] J. C. Wolford and A. S. Hall, Jr., "Second-Acceleration Analyses of Plane Mechanisms," ASME Paper No. 57—A-52.

[7] Professor, City College of New York, New York, N. Y.

The Measurement of Very Low Frictional Torques in Rotating Equipment

M. ZAID
Technical Consultant,
Technik Incorporated, Syosset, N. Y.
Assoc. Mem. ASME

I. S. TOLINS
Principal Engineer,
Ford Instrument Corporation,
Long Island City, N. Y.
P. E., Assoc. Mem. ASME

In the design of high precision, rotating equipment such as air bearings, gyroscopes, etc., it has become possible to reduce the frictional torques to "almost zero" values. The measurement and classification of the sources of these minute torques during rotation is exceedingly difficult, but very important for correct application and further design improvement.

It is the purpose of this paper to present two techniques for the measurement and discrimination of the torques, and an example showing the accuracy of these techniques.

Although these methods were designed specifically for the measurements of low torques, they work equally well for any magnitude and as such present a good evaluation technique. A comparison made with the values obtained by one of the better "direct" methods indicates that the direct method is inadequate.

Introduction

In the design of high precision, rotating equipment it has become possible to achieve "almost zero" frictional torques. The measurement of these torques without influencing them by test apparatus, and the classification of their sources becomes increasingly difficult as these torques are reduced to minute values. For instance, in the design of air bearings for precision instruments it is desirable to minimize the coulomb friction and measure the viscous friction coefficient when the bearing is running at operating speeds. The coulomb friction may come from contacts which are sometimes used to indicate the relative rotational position of the inner and outer cylinder or from small particles and burrs that may exist undetected in the very narrow gaps between the two cylinders. The viscous friction is that component associated with the shear of the air between the two cylinders. The problem of discriminating between viscous and coulomb friction is quite difficult since the viscous drag is essentially constant over a narrow velocity range and behaves very much like the coulomb friction. However, this discrimination is very important in the design of contacts with a minimum of friction or in the inspection of a bearing to determine how "clean" it is.

Several "direct" techniques were attempted in advance of the development of the methods finally presented in this paper. They all failed because of complexity, lack of accuracy, or because the measuring devices affected the experiment. With respect to this latter point, it should be noted that it is desirable to detect torques of the order of 5 dyne-cm or lower. One such technique, which was given considerable attention, requires an experimentally obtained angular displacement-time relationship and from this (or by separate measurement) its first and second derivatives. Writing the dynamical equations of motion and substituting known values of angular displacement, velocity, and acceleration, simultaneous algebraic equations for the evaluation of the viscous and coulomb friction are obtained. However, because this technique uses known values at only two different times to characterize the entire displacement-time curve, it is necessary that they be extremely accurate. In tests run by the authors the first and second derivatives of this curve could not be obtained with the degree of accuracy required for this method. As a consequence, the consideration of various pairs of displacement, velocity, and acceleration readings led to a wide variety of coulomb and viscous torque values and did not allow any single consistent pair of friction values to be selected with accuracy or assurance. Variations of these techniques based on the integrated equations of motion were obtained in the course of this investigation. They were not pursued in great detail because of the reasons to be listed and because it was felt that they still retained the disadvantage of characterizing an entire curve by two or three experimental points, and as such might lead to inaccuracies.

It is the purpose of this paper to present two simple but highly accurate techniques for the evaluation of the coulomb and viscous

Contributed by the Machine Design Division and presented at the Annual Meeting, Atlantic City, N. J., November 29–December 4, 1959, of The American Society of Mechanical Engineers.

Note: Statements and opinions advanced in papers are to be understood as individual expressions of their authors and not those of the Society. Manuscript received at ASME Headquarters, January 6, 1959. Paper No. 59—A-43.

Nomenclature

$c^{\pm}$ = viscous drag coefficient of the system, c^{+} applies when rotation is in the positive direction, c^{-} applies when the rotation is in the negative direction, consistent units

I = mass moment of inertia with respect to the axis of rotation, consistent units

ln = log to base e

$T^{\pm}$ = constant torque component, T^{+} applies when rotation is in positive direction, T^{-} applies when rotation is in negative direction, consistent units

T_f = resisting torque component, positive in direction of positive rotation, consistent units

T_c = coulomb torque component, opposite to the direction of rotation, consistent units

T_v = viscous torque, consistent units

t = time, sec

θ = angular displacement, radians

$\dot{\theta}, \ddot{\theta}$ = angular velocity and acceleration

$\Phi = \int_0^{\theta} \dot{\theta} d\theta$, the area under the $\dot{\theta} - \theta$ curve

$\Delta\Phi$ = area between tangent and actual $\dot{\theta} - \theta$ curve

Subscripts

0 = starting value

F = final value

T = tangent curve

friction. These methods can be classified under the heading of decrement techniques. Along with the well-known log-decrement method associated with viscous damping, they take into account the variation from free rotation conditions. Since they measure the main frictional effect directly, such techniques furnish accurate, convenient, and direct approaches to this problem. The methods of this paper were selected because the associated data are plotted as a straight line or a deviation from a straight line on regular graph paper. This was preferred to the log-decrement approach because the log plot needed for straight line representation of viscous damping tends to mask the difference between small frictional effects and experimental error, so that neither accuracy or interpretation is adequate for cases of very low friction. In addition, the approaches introduce new concepts that could prove useful in the study of dynamic systems with friction elements. Although these methods were originally intended to measure very low frictional torques, they can be used equally well for any magnitude torque and, as such, present useful general techniques for the evaluation and classification of frictional torques.

Theory

General Considerations. Fig. 1 illustrates only the freely rotating cylinder and the torques resisting motion in the system under consideration. The direction of positive angular motion and torques have been drawn on the figure. The inertia force is $I\ddot{\theta}$, the viscous torque (T_v) is $c^{\pm}\dot{\theta}$, and the constant torque component is $T^{\pm}$; the plus superscripts being taken when the angular velocity is positive, and the negative superscripts when it is negative.

Fig. 1

Although many applications occur for which the coefficients associated with the resisting torques (viscous and/or coulomb) are independent of the direction of rotation, no such assumption will be made here. We will assume only that the viscous system coefficient $c^{\pm}$ and the constant torque component $T^{\pm}$ are constant for the speed variation under test and that these quantities have, at most, a change of magnitude with change of direction of rotation. If desired, by a series of tests as described in this paper, the variation of these values with speed can be ascertained.

By evaluating the constant torque component $T^{\pm}$ in each direction, the following interpretation can be given. $T^{\pm}$ is written as the sum or difference of two components, T_f and T_c, both having magnitudes that are independent of the direction of motion. T_f is taken positive in the direction of positive motion and to have a direction which is independent of the direction of motion, and T_c is taken to act so as to oppose the motion of the rotating cylinder. Thus

$$T^{+} = T_f + |T_c| \tag{1}$$

$$T^{-} = T_f - |T_c| \tag{2}$$

For convenience we can visualize the torque component T_f as corresponding to a constant extraneous forcing torque, such as that which induces steady-state drift in an air bearing when $|T_f| > |T_c|$. In this case the uni-directional torque component arises from the effect of surface imperfections on the air flow characteristics. Although these imperfections are unpredictable and their effects presently uncalculable; they do result in a non-symmetrical air flow pattern with a characteristic torque direction. This torque is sometimes referred to as "turbine torque" because of its tendency to produce a "turbining" effect. $|T_c|$ can be thought of as that component of coulomb friction that changes in sign, but not magnitude, as the direction of motion changes. From equations (1) and (2)

$$T_f = \frac{T^{+} + T^{-}}{2} \tag{3}$$

$$|T_c| = \frac{T^{+} - T^{-}}{2} \tag{4}$$

Further, although these torque values may vary slightly with angular position, it is assumed that there is an average value over each revolution that can be used as a measure for the friction coefficient. It is this average value which will be evaluated in this paper.

The experimental method of this paper is to bring the rotating cylinder up to operating speed, then cut the power and allow the coulomb and viscous drag to decay the cylinder speed. Using the measured quantities and techniques indicated in the following sections, the desired values are obtained.

The system equation of free motion is,

$$\ddot{\theta} + \frac{c^{\pm}}{I}\dot{\theta} = -\left(\frac{T^{\pm}}{I}\right) \tag{5}$$

Method I. Taking $\theta = 0$ and $\dot{\theta} = \dot{\theta}_0$ at $t = 0$, the first integral of equation (5) is

$$\frac{\Delta\dot{\theta}}{t} - \frac{c^{\pm}}{I}\left(\frac{\theta}{t}\right) = \frac{T^{\pm}}{I} \tag{6}$$

where the velocity change $\Delta\dot{\theta}$ is $\dot{\theta}_0 - \dot{\theta}$.

Equation (6) expresses a linear relationship between θ/t and $\Delta\dot{\theta}/t$ which can be used as the basis of a test to find $c^{\pm}/I$, $T^{\pm}/c^{\pm}$, and $T^{\pm}/I$. This phase plane representation is shown in Fig. 2, for the case where $\dot{\theta}_0 > 0$, $T^{\pm} = T^{+} > 0$, and $c^{\pm} = c^{+}$. The test requires $\theta - t$ data as well as $\dot{\theta} - t$ data. The latter can either be measured directly or else derived from the $\theta - t$ curve.

The points between P_3 and P_4 will show scatter and should be discarded. The spread occurs in this region because the values of θ, $\Delta\dot{\theta}$, and t are very small and the ratios of θ/t and $\Delta\dot{\theta}/t$ are essentially experimentally indeterminate (0/0) and very sensitive to the slightest inaccuracies. Once sufficiently large values of θ, t, and $\Delta\dot{\theta}$ are obtained this spread will vanish and, even with relatively crude apparatus, the points in the region between P_2 and P_3 will lie on a straight line. P_2 is the point at which the experiment can be considered to terminate either because the velocity has fallen too far below the operating velocity or else has reached zero. The straight line between P_1 and P_2 is an extrapolation.

As is apparent from the phase plane (Fig. 2) or equation (6), the desired value of $c^{\pm}/I$ is obtained as the slope of the $\Delta\dot{\theta}/t - \theta/t$ curve. The value of $T^{\pm}/I$ is the $\Delta\dot{\theta}/t$ intercept, and that of $-T^{+}/c^{+}$ is the θ/t intercept. Care must be taken to use the proper algebraic sign with the value of these intercepts.

With the use of equation (6), the discussion is generalized to include all $\dot{\theta}_0$ and $T^{\pm}$ combinations. The heavy lines in Fig. 3 illustrate the phase plane paths for each possible $\dot{\theta}_0$, $T^{\pm}$ combination. The direction of generation of these paths with increasing time is shown on the figure. These directions are obtained from consideration of the solution to equation (5) which indicates that $|\Delta\dot{\theta}/t|$ monotonically decreases with increasing time.

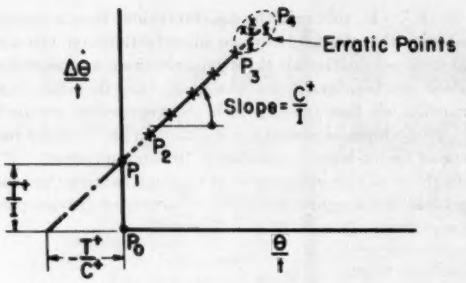


Fig. 2

Paths "A" and "C" are traced when the resisting torque is opposite to the direction of motion. In the event that $|T_e| > |T_f|$, the motion will stop in a finite time. However, when $|T_e| < |T_f|$, the motion will stop in a finite time and will start in the reverse direction. In that event the path will end at P_1 on "A" and start at P_2 on "B" (or correspondingly on "C" and "D"); the initial conditions for the reversed motion along path "B" (or "D") being that at point P_2, $t = 0$, $\theta = 0$, $\dot{\theta}_0 = 0$.

Paths "B" and "D" are traced when $|T_f| > |T_e|$ and $T^{\pm}$ is in the direction of the motion. In that event a steady-state velocity ($|\dot{\theta}| = |T^{\pm}|/c^{\pm}$) will be reached at infinite time (point P_4). Depending on whether the magnitude of the starting velocity ($\dot{\theta}_0$) is greater or smaller than the magnitude of the steady-state velocity ($|T^{\pm}|/c^{\pm}$) the rotor speed will monotonically decrease or increase until the steady-state condition is attained. Thus when $|\dot{\theta}_0| < |T^{\pm}|/c^{\pm}$ the path will be from P_3 to P_4 and when $|\dot{\theta}_0| > |T^{\pm}|/c^{\pm}$ the path will be from P_3 to P_4.

In the special event that paths "B" (or "D") are applicable to a design and the running speed under consideration is the steady-state speed, one test will not suffice. In this case only the point P_4 which specifies $T^{\pm}/c^{\pm}$ is obtained, but the path slope or the $\Delta\dot{\theta}/t$ intercept is lacking. A second start at above or below this steady-state speed will be necessary for the complete evaluation of all the desired parameters.

Experimental results for both methods I and II will be given after the next section.

Method II. Taking the same initial conditions as in method I, a first integral of equation (5) can be shown to be

$$\frac{\dot{\theta}^2 - \dot{\theta}_0^2}{2} + \frac{c^{\pm}}{I}\Phi = -\frac{T^{\pm}}{I}\theta \tag{7}$$

where

$$\Phi = \int_0^{\theta} \dot{\theta} d\theta \tag{8}$$

As for the previous method, a detailed discussion will now be

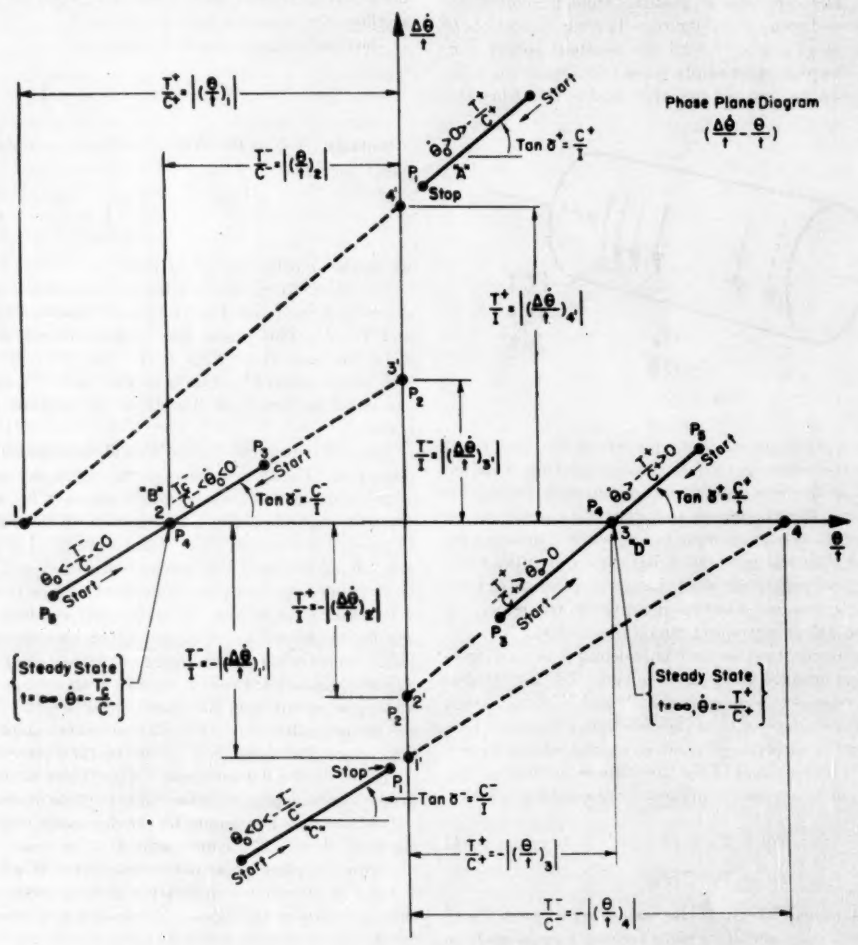


Fig. 3

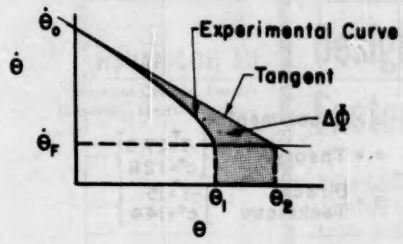


Fig. 4

undertaken for the case of $\dot\theta_0 > 0$ and T^+ resisting the direction of motion. The results will then be generalized, leaving some of the details for the interested reader.

For the afore-mentioned case, as will be explained, it can be shown that the $\dot\theta - \theta$ curve is concave downward as shown in Fig. 4.

The direction of curvature can be determined by considering $\dot\theta$ and its second derivative with respect to θ,

$$\dot\theta = \dot\theta_0 - \frac{T^\pm}{c^\pm}\left(1 - e^{-\frac{c^\pm}{I}t}\right) \tag{9}$$

$$\frac{d^2\dot\theta}{d\theta^2} = -\left(\frac{c^\pm}{I\dot\theta}\right)^2 \frac{\left(\frac{T^\pm}{c^\pm}\right)}{(\dot\theta)}\left(\dot\theta + \frac{T^\pm}{c^\pm}\right) \tag{10}$$

For the case under consideration, $\dot\theta_0 > 0 > -T^+/c^+$, which with the use of equation (9) leads to $\dot\theta \geq 0$ for all t; $\dot\theta = 0$ being the end of the test. From equation (10) there results $d^2\dot\theta/d\theta^2 < 0$, which is the previously stated conclusion on the concavity of the $\dot\theta - \theta$ curve.

Taking the first derivative of $\dot\theta$ with respect to θ and setting $\dot\theta = \dot\theta_0$, the slope of the tangent line at $\theta = 0$ is given by

$$\left(\frac{d\dot\theta}{d\theta}\right)_{\theta=0} = -\frac{c^\pm}{I} - \frac{T^\pm}{I\dot\theta_0} \tag{11}$$

With reference to Fig. 4, let us define $\Delta\Phi$ as the area between the tangent and experimental curve limited by $\dot\theta_0$ and $\dot\theta_F$, plus the area contained in the small rectangle below $\dot\theta_F$ which is bounded by θ_1 and θ_2. Once the curve and its tangent are drawn, this area is easily obtained by means of a planimeter, square counting, or any other suitable technique.

The area bounded by the tangent curve and the vertical line at $\theta = \theta_2$ is obtained as

$$\Phi_T = \left(\frac{\dot\theta_0^2 - \dot\theta_F^2}{2}\right)\frac{1}{\left(-\frac{d\dot\theta}{d\theta}\right)_{\theta=0}} \tag{12}$$

The value of $(-d\dot\theta/d\theta)_{\theta=0}$ is a positive number because the slope is negative (see Fig. 4).

Using equations (7) and (12) to form $\Delta\Phi$, and equation (11) to eliminate $c^\pm$, the equation for the evaluation of $T^\pm$ becomes,

$$\frac{T^\pm}{I} = \frac{\left(-\frac{d\dot\theta}{d\theta}\right)_{\theta=0}\Delta\Phi}{\theta_1 + \frac{\Delta\Phi}{\dot\theta_0} - \frac{\dot\theta_0^2 - \dot\theta_F^2}{2\dot\theta_0\left(-\frac{d\dot\theta}{d\theta}\right)_{\theta=0}}} \tag{13}$$

$c^\pm/I$ is evaluated from equation (11), once $T^\pm/I$ is determined.

The generalization to include all $\dot\theta_0$ and $T^\pm$ combinations can now be represented with the aid of equations (9) and (10). The heavy lines on Fig. 5 illustrate the phase plane paths, with the direction of increasing time being indicated by arrow.

The description of the generation of each phase plane path is identical to that given in conjunction with method I, Fig. 3. The only modification being that, for the case where the motion reverses itself, curve "B" (and correspondingly "D") will be

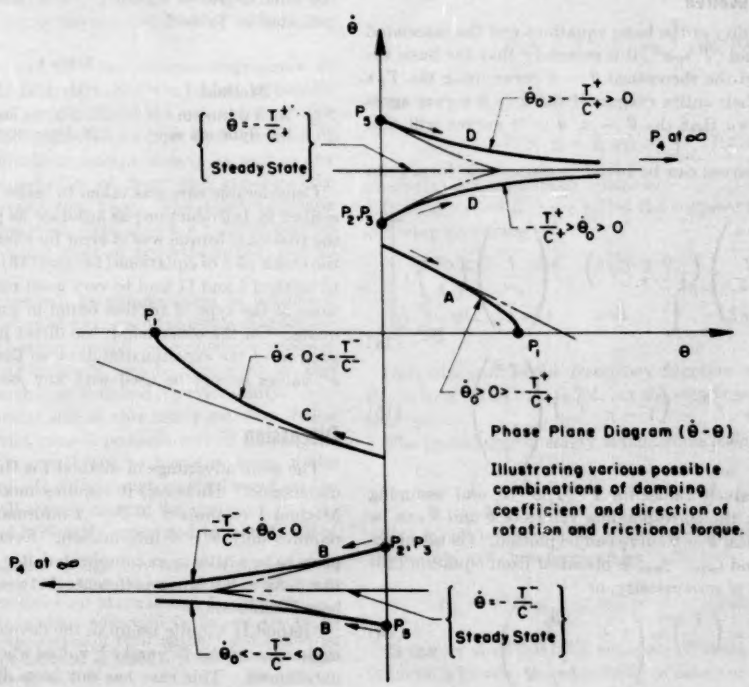


Fig. 5

Phase Plane Diagram ($\dot\theta$-θ)

Illustrating various possible combinations of damping coefficient and direction of rotation and frictional torque.

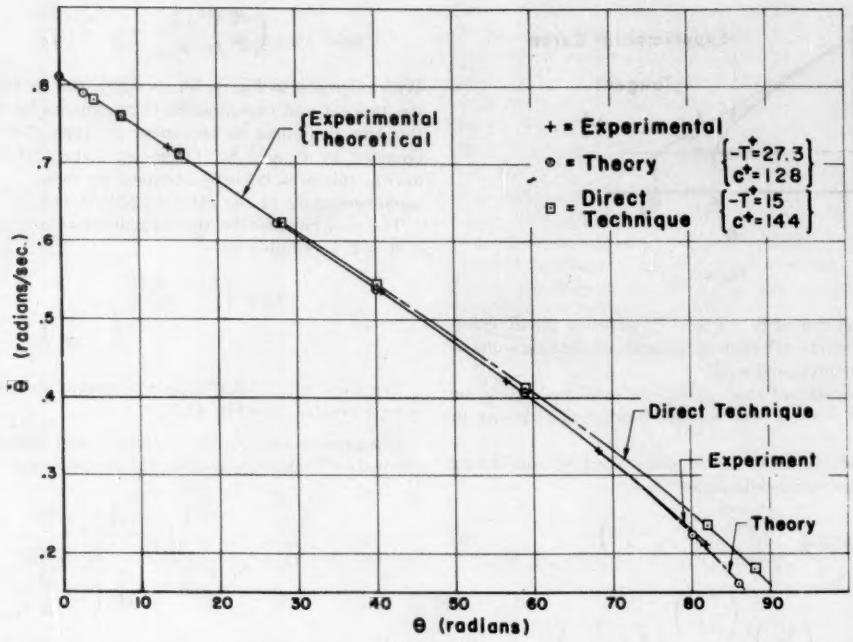


Fig. 6

translated parallel to itself until P_2 passes through the origin since the initial conditions for the second part of the motion are $t = \theta = \dot{\theta} = 0$.

Should it be desired, a formula for the evaluation of $T^{\pm}$ for the case of $|T_f| > |T_c|$ can be obtained in much the same manner as was used for equation (13).

Experimental Verification

To check the validity of the basic equations and the associated pair of derived values ($T^{\pm}, c^{\pm}$) it is necessary that the basic experimental data and the theoretical $\dot{\theta} - \theta$ curve using the T, c values agree over their entire range. If the $\dot{\theta} - \theta$ curves agree well, it can be shown that the $\theta - t$, $\dot{\theta} - t$ curves will also agree well.

The theoretical curves can be found in parametric form from equation (5) to be

$$\theta = \frac{I}{c}\left[\left(\dot{\theta}_0 + \frac{\frac{T^{\pm}}{I}}{\frac{c^{\pm}}{I}}\right)\left(1 - e^{-\frac{c^{\pm}}{I}t}\right) - \left(\frac{\frac{T^{\pm}}{I}}{\frac{c^{\pm}}{I}}\right)\left(\frac{c^{\pm}}{I}\right)t\right] \tag{14}$$

$$\dot{\theta} = \dot{\theta}_0 e^{-\frac{c^{\pm}}{I}t} - \left(1 - e^{-\frac{c^{\pm}}{I}t}\right)\left(\frac{\frac{T^{\pm}}{I}}{\frac{c^{\pm}}{I}}\right) \tag{15}$$

Substituting the derived values for $T^{\pm}/I$, $c^{\pm}/I$, and assuming different values of t, the corresponding values of $\dot{\theta}$ and θ can be found and a theoretical $\dot{\theta} - \theta$ curve can be plotted. t is taken between values of 0 and t_{max}. t_{max} is obtained from equation (15) using the conditions of zero velocity, or

$$t_{max} = \left(\frac{I}{c^{\pm}}\right)\ln\left(1 + \frac{\dot{\theta}_0}{\frac{T^{\pm}}{I}\Big/\frac{c^{\pm}}{I}}\right) \tag{16}$$

As an example of the methods that have been described, the curve shown in Fig. 6 was obtained from an air bearing. Since values for t were not available from the test data, they were computed from equation (14) and the best $c^{\pm}$, $T^{\pm}$ values. The computed t together with the experimental θ and $\dot{\theta}$ values were used to obtain the results listed under method I. However, in later work t was obtained experimentally and the results showed the same degree of accuracy. The calculated T^{+}, c^{+} values are indicated in Table 1.

Table 1

	Method I	Method II	Direct method
T^{+}	26.5 dyne cm	27.3 dyne cm	15 dyne cm
c^{+}	136 dyne cm sec	128 dyne cm sec	144 dyne cm sec

Considerable care was taken to make the direct technique (described in Introduction) as accurate as possible, but nevertheless the frictional torque was in error by a factor of 2. It is seen from the check plot of equations (14) and (15) on Fig. 6 that the results to method I and II lead to very good results so that the assumptions of the type of friction found in such an air bearing is adequate. On the other hand, the direct method does not lead to a good fit of the experimental data, so that the associated T^{+} and c^{+} values cannot be used with any assurance.

Discussion

The main advantage of method I is that it is simple to use and understand. However, it requires more data than method II. Method I requires $\dot{\theta} - \theta - t$ information whereas method II requires only $\dot{\theta} - \theta$ information. Even though method II appears to be a little more complicated, it is quite easy to use in practice. As a result no preference between the two methods has been established.

Method II is quite useful in the development of theory for the evaluation of the $T^{\pm}$ and $c^{\pm}$ values when the rotating cylinder is unbalanced. This case has not been discussed in the paper.

HSUAN-LOH SU
Imperial College, London University, London, England

Design by Quantitative Factor of Safety

The traditional safety factor has a qualitative meaning only. However, it can be quantitatively interpreted through the possible variations of the components in a design as well as the probability of safety. The proposed design method based on such a safety factor gives not only a more rational but also a more economic design than the conventional methods.

Introduction

THE DETERMINATION of an adequate "safety" factor in structure and machine design has always been of paramount importance. The most probable factor of safety is thought to depend on the accuracy of the knowledge of the loading likely to act on the material, the homogeneity of material, the interaction between the external loading and the internal response to such disturbance, and so forth. However, after such a correct diagnosis, no cure was thought of by engineers. In the normal design convention, this factor is only arbitrarily determined and its value can vary from 1.5 for most members in aerostructures, 2.0 for the pressurized cabin in jet planes, 2.5 for bridges, 3.0 for foundations, up to 6.0 for boilers.

There are two principal aspects of any engineering design, the physical and the statistical aspect. Since the appearance of many modern physical concepts of the macroscopic mechanics, most research workers in this field have been heavily occupied and have forgotten that all these are only one side of the design problem. The other side signified by the safety factor is hardly considered, and little attention is paid to the fact that all the variables involved in any design are of a stochastic nature, which is completely ignored in almost every design convention. This explains why the safety factor, having lost its true essence, remains a factor of ignorance.

Broadly speaking, there can be three different approaches to the conception of safety. The first also the oldest approach considers the prevention of accidents and the risk of human life as the aim of a design. The second results from the realization of the limited resources available to human society as well as the comparably short life expectation of man, and consequently economy plays an important role here. The third approach covers all the other considerations. The object of a design may be the military efficiency or some kind of advertisements, and so forth.

The first two approaches can be compromised to achieve a desired serviceability, which can be expressed in terms of probability based upon the risk of human life and the economic consideration. The term safety, although it will be preserved in this article, actually means such compromised "serviceability."

However, the philosophical side of this safety problem, being so complicated and involved, cannot be dealt with in this article. Valuable reference on this aspect can be found in most of the papers in the Bibliography. It is the design method based on the statistical concept that will occupy most of the space.

Since the design method is directly connected with methods of analysis as well as other relevant fundamental investigation, the statistical approach to be suggested will no doubt influence the design philosophy. It is therefore desirable to discuss the possible impact of such an approach on the future design technique and on the physical investigation. There will be two sections devoted to this purpose.

In order to illustrate the technique of application rather than the underlying theory of statistics, readers will be assumed to possess the elementary knowledge of necessary mathematics. Further reference on statistics can be obtained from any standard textbook. It should be emphasized that this article is only an introduction to the statistical concept of the design method, and by no means a thorough exploration of such a vast field. On the other hand, the practical technique developed here is perfectly general and applicable to all design problems.

Notations used in this article will be explained on their first appearance except those in the examples, which are essentially the same as Timoshenko's system.

Probability of Safety

The probability of the occurrence of a variable x in the range $x_1 \leqq x \leqq x_2$ can be defined as follows:

$$P(x_1 \leqq x \leqq x_2) = \sum_{x_1}^{x_2} \phi(x)\Delta x \tag{1}$$

where $\phi(x)$ is a step function, or

$$P(x_1 \leqq x \leqq x_2) = \int_{x_1}^{x_2} f(x)dx \tag{2}$$

where $f(x)$ is a continuous function.

Both $\phi(x)$ and $f(x)$ are called the frequency function with the following properties

(i) $\phi(x) \geqq 0$ and $f(x) \geqq 0$;

(ii) $\sum_{-\infty}^{+\infty} \phi(x)\Delta x = 1$ and $\int_{-\infty}^{+\infty} f(x)dx = 1.$

Only the continuous frequency function will be discussed in this article. The case involving the step function can be treated analogously.

The probability of safety is defined as follows:

$$P_s = P(-\infty < x \leqq \alpha) = \int_{-\infty}^{\alpha} f(x)dx \tag{3}$$

where α will be referred to as the safety limit. The probability of unsafety will then be given by

$$P_u = P(\alpha \leqq x < \infty) = \int_{\alpha}^{\infty} f(x)dx = 1 - P_s$$

It can be observed from equation (3) that once the frequency function is known, the probability of safety or unsafety is uniquely

Contributed by the Machine Design Division and presented at the Annual Meeting, Atlantic City, N. J., November 29–December 4, 1959, of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS.
NOTE: Statements and opinions advanced in papers are to be understood as individual expressions of their authors and not those of the Society. Manuscript received at ASME Headquarters, October 21, 1958. Paper No. 59—A-48.

determined by the safety limit α. When the frequency function is not known at all, the lower bound of the probability of safety corresponding to a given safety limit can be obtained from Tchebycheff's inequality, which may be expressed as follows:

$$P_s \geqq 1 - \frac{1}{\alpha^2} \quad \text{or} \quad P_u \leqq \frac{1}{\alpha^2} \tag{4}$$

Some numerical values given by equation (4) are tabulated in Table 1, in which results based upon the normal frequency functions are also included for comparison.

In an actual design problem, the first step is to determine the safety limit α, which is related to a definite P_s or P_u. Then, by use of α, it is possible to represent the expected safety in terms of the safety factor, which is also based upon the variabilities of the component quantities. By such manipulation, the factor safety can possess a quantitative significance.

Concept of Safety Factor

The factor of safety β is conventionally defined as a ratio of two quantities, say, R and Q, each being associated with the unsafe or unserviceable condition and with the safe or serviceable condition.

$$\beta = R/Q \tag{5}$$

The most familiar ratio is the yield stress over the working stress.

Since R and Q are both statistical quantities, i.e., obtained through a statistical treatment, the factor of safety being the ratio of them, should also be dealt with statistically. Let σ_β, σ_R, and σ_Q denote the standard deviation of β, R, and Q. Then, by the variance law in statistics,

$$\sigma_\beta^2 = \left(\frac{\partial\beta}{\partial R}\right)^2 \sigma_R^2 + \left(\frac{\partial\beta}{\partial Q}\right)^2 \sigma_Q^2 + 2\rho_{RQ}\sigma_R\sigma_Q \frac{\partial\beta}{\partial R}\frac{\partial\beta}{\partial Q} \tag{6}$$

where R and Q are assumed to have some correlation, and the correlation coefficient is ρ_{RQ}. If R and Q are uncorrelated or independent, $\rho_{RQ} = 0$.

From equation (6), the following relation among the variabilities can be obtained

$$\gamma_\beta^2 = \left(\frac{R}{\beta}\frac{\partial\beta}{\partial R}\right)^2 \gamma_R^2 + \left(\frac{Q}{\beta}\frac{\partial\beta}{\partial Q}\right)^2 \gamma_Q^2 + 2\rho_{RQ}\gamma_R\gamma_Q \left(\frac{RQ}{\beta^2}\frac{\partial\beta}{\partial R}\frac{\partial\beta}{\partial Q}\right) \tag{7}$$

where γ_x denotes the variability of x and $x = \beta$, R, Q, and so forth.

In almost all of the practical cases, R and Q can be expressed in terms of several uncorrelated variables. The generalized form of Equation (6) and (7) will be

$$\sigma_\beta^2 = \left(\frac{\partial\beta}{\partial A}\right)^2 \sigma_A^2 + \left(\frac{\partial\beta}{\partial B}\right)^2 \sigma_B^2 + \ldots + \left(\frac{\partial\beta}{\partial N}\right)^2 \sigma_N^2 \tag{8}$$

$$\gamma_\beta^2 = \left(\frac{A}{\beta}\frac{\partial\beta}{\partial A}\right)^2 \gamma_A^2 + \left(\frac{B}{\beta}\frac{\partial\beta}{\partial B}\right) \gamma_B^2 + \ldots + \left(\frac{N}{\beta}\frac{\partial\beta}{\partial N}\right)^2 \gamma_N^2 \tag{9}$$

It should be noted that, in these equations, $\beta, A, B, \ldots, N$, and so forth, are the means of the variables and not the variables themselves.

When the variability of the safety factor γ_β is known, the required value of the safety factor for a particular design is defined by

$$\beta = 1 + \alpha\gamma_\beta \tag{10}$$

where α is the safety limit. In the case that the standard deviation of the safety factor σ_β is known, the required safety factor will be given by

$$\beta = \frac{1}{2} + \frac{1}{2}\sqrt{1 + 4\alpha\sigma_\beta} \tag{11}$$

Equation (10) manifests the meaning of the safety factor, which is directly related to the probability of safety through α and to the possible variation of the performance of a design through γ_β. To reduce the factor of safety of a design, it is necessary either to reduce the possible variation of the performance through methods of quality control, and so forth or to accept a lower safety limit corresponding to a reduced probability of safety. In other words, γ_β can be physically controlled through accurate theoretical prediction, good workmanship, the use of high-quality materials, and so forth, whereas the determination of α will lead to a question on philosophical basis of the design.

The difference between equation (5) and (10) [or equation (11), which is another form of equation (10)] should be noted: Equation (5) defines the general relation between variables while equation (10) gives a particular value of β corresponding to a give nprobability of safety as well as the probable variations in the performance of the particular design. Since equation (5) will not be used to refer to the means henceforth, no distinction will be made between the symbols representing the mean β and the particular value β. In general, Greek letters always denote the required particular value.

In the derivation given, only the statistical parameters of the population are involved. Such parameters as means and variances are also supposed to have been obtained in the following examples. However, the difference between the sample variance and the population variance should be borne in mind, and it is merely for the sake of simplicity in the explanation of the method that the population becomes our sole interest at present. As the population parameters can be easily transformed into the sample parameters, such restriction imposes hardly any limitation on the application of the method developed here.

Safety Limit and Probability of Safety

The importance of σ_β and γ_β has been revealed by equation (10) and (11). The accuracy of the prediction given by equation (10) or (11) depends entirely upon the reliability of σ_β and γ_β, which are derived from the corresponding values of the components. Therefore the problem will then become whether these γ-values are available and reliable.

The value of the variability of the safety factor given by equation (7) or (9) is merely an arbitrarily chosen quantity to signify the accuracy and should not be confused with the actual probable variation which is linked with the safety limit to correspond to a definite likelihood of such variation. From Table 1, the lower bound of the probability of unsafety at $\alpha = 5.0$ is 4 per cent, and the values for P_u following a certain distribution are very small. The value of 5 for the safety limit is therefore very much on the conservative side. The selection of an α-value suitable for a particular design will inevitably involve the philosophy and the policy on the question of safety, and is beyond the scope of the present article. Nevertheless, since this α-value is the essence of the proposed method, it is desirable to review some of the existing statistics.

In the "Return of Shipping Casualties and Deaths, 1957," published by the Ministry of Transport and Civil Aviation of Britain, the gross tonnage of shipping lost from 1947 to 1957 was expressed as a percentage of the corresponding total tonnage, and the annual average is 0.21 per cent. If the expectation of life of a ship is 30 years, then the probability of its being lost during the

service period will be approximately 6.3 per cent for an average tonnage ship. This also means that approximately six ships out of every hundred in use would be lost owing to some accidents.

In the "Road Accidents," published in 1956 by the aforementioned Ministry, it can be observed that the probability of being killed on the roads in Britain is around 0.01 per cent every year and that of being injured or killed is around 0.54 per cent every year. This statistics must have excluded some of the minor unreported accidents, and the actual probability of being involved in a road accident must be higher. If the expectation of life for people living in Britain is, say, 70 years, then the probability for any person being killed in a road accident during his life will be approximately 0.7 per cent and the probability of being injured will be 37.8 per cent; i.e., one out of every three people will suffer some injury of various seriousness on the roads during his life.

The value of α used in the examples will be 2.5, which corresponds to a "normal" probability of unsafety of 0.62 per cent, or about the same as the probability of unsafety on British roads, also about one tenth of that for ships registered in Britain. It should be noted that for a certain class of designs, a definite range of α-values should be specified to allow for a certain amount of flexibility; and the safety factor is not necessarily unique for a given class of designs. The reason is that α is completely defined by the probability of safety, whereas β also depends on the quality of material, workmanship, and so forth.

The inaccuracy of a physical theory, which is not explicitly expressible in a physical equation, can also be incorporated in the safety factor. Suppose the value associated with the unsafe condition is not sufficiently good to approximate the actual condition of unsafety due to some inaccurate assumptions. The true value of R of the prototype may be, say, either 20 per cent higher or 20 per cent lower than the value of R indicated by the theory. If the standard deviation of such variation is, say, 10 per cent, then the constant unity in front of R is said to have a variability of 0.10 which can be inserted into equation (8) or (9) in the practical calculation.

Attention must be paid to the fact that in such a case, R cannot be substituted by $1.2R$ as frequently done by designers. This 1.2 has no physical significance, but is supposed to indicate the statistical characteristics of R, which should obviously be treated statistically. Furthermore, such departure of the theory from reality should not be confused with variation of R itself as a consequence of accidental errors in the experiments and the nonconformity of the test model with the prototype.

Illustration

The problem chosen for illustration is extremely simple so that the principle of the proposed method rather than the details can be easily grasped. By the same principle, more complicated cases can easily be tackled.

The procedure of calculation can be summarized as follows:

(1) Write down the physical equations governing R and Q in terms of their component quantities. R is a function of variables associated with the limit of the serviceable condition, beyond which the structure "fails." Q is another function of variables associated with the expected working condition.

(2) Write down the equation defining the safety factor.

$$\beta = \frac{R}{Q}$$

(3) Work out the coefficients required for calculation of σ_β and γ_β; i.e., to establish the σ or γ equation.

(4) Compute β, if necessary by trial and error, to correspond to a specified probability of safety signified by α.

(5) The criterion for a "safe" design is then

$$R \geq \beta Q$$

The example is a simply supported rectangular steel plate. The solution has been given by Timoshenko in his "Theory of Plates and Shell." Other given data are: Uniformly distributed load of 80 lb $\pm$ 20 lb, yield stress $y = 35$ kip/sq in., and the size of the plate is 2 ft 6 in. $\times$ 10 ft 0 in. The physical equation governing its bending behavior is

$$0.1235\, wa^2 = M_{max} = \frac{yh^2}{6}$$

In a conventional design, therefore, the thickness will be

$$h^2 = \frac{1}{20} \times 0.1235 \times 0.1 \times 2.5^2 \times 12 \times 6 = 0.272,$$

$$h = 0.527 \text{ in}$$

The equation of safety factor is

$$\beta = \frac{yh^2}{0.741wa^2}$$

The γ-equation is

$$\gamma_\beta^2 = \left(\frac{y}{\beta}\cdot\frac{\partial\beta}{\partial y}\right)^2\gamma_y^2 + \left(\frac{w}{\beta}\cdot\frac{\partial\beta}{\partial w}\right)^2\gamma_w^2 + \left(\frac{a}{\beta}\cdot\frac{\partial\beta}{\partial a}\right)^2\gamma_a^2 + \left(\frac{h}{\beta}\cdot\frac{\partial\beta}{\partial h}\right)^2\gamma_h^2$$

where

$$\left(\frac{y}{\beta}\cdot\frac{\partial\beta}{\partial y}\right)^2 = \left(\frac{w}{\beta}\cdot\frac{\partial\beta}{\partial w}\right)^2 = 1$$

$$\left(\frac{a}{\beta}\cdot\frac{\partial\beta}{\partial a}\right)^2 = \left(\frac{h}{\beta}\cdot\frac{\partial\beta}{\partial h}\right)^2 = 4$$

Therefore,

$$\gamma_\beta^2 = \gamma_y^2 + \gamma_w^2 + 4\gamma_a^2 + 4\gamma_h^2$$

(It should be pointed out that, in general, γ-equation will be more complicated.)

Suppose the precision of the thickness is much higher than all other quantities involved in the design, $\gamma_h = 0$. Assume that there may be some error in tailoring the plate, take $\gamma_a = 0.05$. Then, with $\gamma_h = 0.20$,

$$\gamma_\beta^2 = (0.2)^2 + (0.25)^2 + 4(0.05)^2 = 0.1125, \quad \gamma_\beta = 0.336$$

Use $\alpha = 2.5$,

$$\beta = 1 + 2.5 \times 0.336 = 1.84$$

Hence

$$h^2 = \frac{1.84}{35}\, 0.741 \times 0.08 \times 2.5^2 \times 12 = 0.234, \quad h = 0.485 \text{ in.}$$

If $^1/_2$-in. plate is used, its actual probability of safety can be estimated as follows:

$$\beta = \frac{0.25}{0.234} \times 1.84 = 1.965 \qquad \alpha = \frac{0.965}{0.336} = 2.87$$

From Table 1, P_s following the normal distribution is interpolated and found to be 0.0024. So $P_s = 0.9976$.

The safety factor in this calculation is based on the yielding limitation (but not plastic collapse limitation), similar procedure can be used for calculation based on the deflection limitation.

Suppose the damaging deflection of the plate is 2 in. The physical equation is

Table 1 Safety limit versus probability of safety and unsafety

	Tchebycheff's theorem		Normal probability	
α	P_s	P_u	P_s	P_u
1.0	0.5	0.5	0.8413	0.1587
1.5	0.778	0.222	0.9332	0.0668
2.0	0.875	0.125	0.97725	0.02275
2.5	0.920	0.080	0.99379	0.00621
3.0	0.944	0.056	0.99865	0.00135
3.5	0.959	0.041	0.9997674	0.0002326
4.0	0.969	0.031	0.99996833	0.00003167
5.0	0.980	0.020	0.9999997133	0.0000002867

$$\Delta = 0.14 \frac{wa^4}{Eh^3}$$

The safety factor equation is

$$\beta = \frac{Eh^3\Delta}{0.14wa^4}$$

The γ-coefficients are

$$\left(\frac{\Delta}{\beta} \cdot \frac{\partial\beta}{\partial\Delta}\right)^2 = \left(\frac{E}{\beta} \cdot \frac{\partial\beta}{\gamma E}\right)^2 = \left(\frac{w}{\beta} \cdot \frac{\partial\beta}{\gamma E}\right)^2 = 1,$$

$$\left(\frac{h}{\beta} \cdot \frac{\partial\beta}{\partial h}\right)^2 = 9, \qquad \left(\frac{a}{\beta} \cdot \frac{\partial\beta}{\partial a}\right)^2 = 16.$$

Therefore the γ-equation is

$$\gamma_\beta^2 = \gamma_\Delta^2 + \gamma_E^2 + \gamma_w^2 + 9\gamma_h^2 + 16\gamma_a^2$$

Assume that the variability in measuring the damaging deflection is 25 per cent; i.e., $\gamma_\Delta = 0.25$; and the other variabilities are

$$\gamma_E = 0.10, \qquad \gamma_h = 0.01, \qquad \gamma_a = 0.05, \qquad \gamma_w = 0.25$$

Then,

$$\gamma_\beta^2 = (0.25)^2 + (0.25)^2 + (0.10)^2 + 9(0.01)^2 + 16(0.05)^2 = 0.1754$$

$$\gamma_\beta = 0.420, \qquad \beta = 1 + 2.5 \times 0.42 = 2.05$$

Therefore,

$$h^3 = \frac{2.05 \times 0.14 \times 0.08 \times (30)^4}{30 \times 10^3 \times 2} = 0.310$$

$$h = 0.678 \text{ in.}$$

It is quite possible that some other effects which are not represented in the physical equations, may influence the variability γ_β considerably. Unless a quantitative evaluation of such effects is possible and reliable, engineer's intuition and experience must come into play. For example, if the plate is made up by two plates welded together along the long dimension, the workmanship must be considered, and the variability of worksmanship should be inserted in the γ-equation.

It is also possible to obtain the factor of safety against buckling by the foregoing procedure. The physical equation governing the stability of the plate is

$$f = \frac{\pi^2 Eh^2}{3a^2(1 - \nu^2)} = 31 \text{ k/sq in. } (h = {}^1/_2 \text{ in., } \nu = 0.3)$$

The safety factor equation is

$$\beta = \frac{\pi^2 Eh^2}{2.73a^2 f}$$

The γ-equation is

$$\gamma_\beta^2 = \gamma_E^2 + \gamma_f^2 + 4\gamma_h^2 + 4\gamma_a^2 + \text{allowance}$$

Take $\gamma_E = \gamma_f = 0.10$, $\gamma_a = 0.05$, $\gamma_h = 0.01$ and no allowance for other effects. Then

$$\gamma_\beta^2 = 0.0304, \qquad \gamma_\beta = 0.175$$

$$\beta = 1 + 2.5 \times 0.175 = 1.44$$

For the $^1/_2$-in. plate, the allowable compressive stress is $^{31}/_{1.44}$ = 21.5 k/sq in. or the allowable compressible force is 21.5 × $^1/_2$ × 12 = 123 k/ft width.

Why Plastic Method of Design

The elastic method of design has been repeatedly denounced since the end of the last war, when a plastic theory on continuum was introduced by Prager and Hill, and a similar one on discrete structures by Baker.

According to the plasticians, a material will yield and flow due to the mechanism occurring under the yield stress. The working stress is then defined as the ratio of the yield stress to an arbitrarily assigned safety factor. Some plasticians work in the reverse direction. They obtain the safety factor from the working stress (or load) and the yield stress (or load) and claim that such a factor is a "true" one.

It is true that a material will yield at its yield stress as this is a pure physical phenomenon. The safety factor on the other hand possesses clear statistical characteristics and is not a physical parameter at all, and hence cannot be obtained through any physical process such as yielding tests.

Of course, the safety factor elaborated in such a manner can be considered as a sheer artificial link between an analysis and a design. However, with such a safety factor embodied in the design, no phenomenon similar to the yielding and flow mechanism will appear under the working condition of the material during its service life. This points out the fact that such a design has little bearing on its actual expected performance.

Obviously it is possible, by the statistical approach, to obtain a design which will yield or will be on the brink of yielding. Only in this case, the reality can be approximated by the plastic theory, and the plastic method of design is acceptable. This argument, however, does not rule out the possibility of existence of the plastic theory as a branch of applied mechanics, but leads to the conclusion that the plastic method can be a sound basis of analysis and design only when the actual plastic state is expected to be attained in the design, such as in wire drawing, metal cutting, and so forth.

It is interesting to note that, although engineers are entrusted to design an object which will work, they are invariably absorbed in the investigation of a similar object which will fail or yield. These two conditions, working and the failure condition, are generally not complementary. In other words, the condition at which a failure occurs may be far beyond the limit at which an object is considered useless; therefore these conditions may be irrelevant. It is apparently more appropriate to select the limit of serviceability than to use the fictitious failure as the design criterion.

Why Destructive Tests

The destructive tests on material have been traditionally performed by engineers in order to investigate the so-called ultimate strength as well as the mechanical behavior immediately before such an ultimate state. The dynamic and the time-dependent properties of materials are still not fully understood at the present and may require such tests to establish the physical concept.

So far as the static properties are concerned, the destructive tests are not necessarily the most suitable routine for a design.

As mentioned heretofore, the purpose of such test is to obtain some kind of "ultimate" data to be used in a design. These tests are therefore directly connected with the present design philosophy which exacts such information. Traditionally, engineers use the safety factor, which is determined by no other criterion than experience, to prevent the designed object from reaching the ultimate state. However, when the ultimate data have been, say, halved, i.e., with a safety factor of two, they can no longer represent the ultimate state, from which they were derived. In fact since only the "halved" values are useful in the design it is quite logical to obtain them directly in the routine tests without doing any destructive effects to the specimens. The foregoing discussions of course apply only to a design which is expected to behave safely during its service life.

Many destructive tests can, as a consequence of the new approach to design problems, be easily modified to become nondestructive ones. The only difference between them is on the application of test loading. In the usual tests the ultimate loads are applied finally. Now since such ultimate data are no longer a necessity in every case, only a test "working" load with due allowance for possible variation is sufficient to furnish information about the future safe behavior of a material.

These nondamaging tests not only can be used in the design stage, but also be employed for checking the designed object after it comes into being. Such checking is possible because all the variations other then the variability of applied loads have been taken into account and liquidated in the design. The partial safety factor, i.e., the safety factor without any allowance for loading variation, is therefore a known quantity corresponding to a given probability of safety. Such tests should confirm the design assumptions and can provide valuable data for the investigation of the probability of safety.

Concluding Remarks

(1) The normal practice of the material manufacturers at present is merely to supply information on the mean values of the required quantities. It is evident from the examples that the variabilities of the required quantities are equally important, and should also be provided by the manufacturers. Furthermore, a material of good quality should be signified not only by its high or low values of a certain properties but also by the smallness of its standard deviation.

(2) Once the variabilities or the standard deviations are available an object can be designed to correspond to a specific probability of safety. On the other hand, with sufficient knowledge of the standard deviation and the variabilities, the probability of safety of the existing design can be assessed and its service life estimated.

(3) The proposed definition of the safety factor indicates its true content clearly. An important feature of such definition is the inclusion of the statistical concept, which enables the suggested method of design to give a more economic result than the conventional method with same limiting conditions.

(4) Either plastic or the elastic state can be chosen as the limiting condition; i.e., the limit of the serviceable condition. Since most of the designed objects are expected to behave safely and not to "yield" during their service life, the plastic method of design seems to have been over-glorified and can in fact only play a role in fail-safe design or wherever a plastic flow is expected. The factor of safety offered by the plastic method is only a factor of ignorance, to which hardly any scientific inference can be attached.

(5) If in certain complicated cases, the vigorous stress analysis is prohibitive, the alternative is to analyze the object in question by a simple method and to introduce a suitable variability due to the inaccuracy to be included in the safety factor. A designer can thus have full control over the design through the flexible use of the safety limit, variabilities, and the safety factor, all of which can be controlled and adjusted to match the role any component plays in the whole assembly.

Bibliography

1 A. H. Chilver, "Some Problems of Structural Safety," *British Welding Journal*, vol. 2, no. 8, 1955.

2 D. B. Dorey and W. R. Schriver, "Philosophy on Loading Tests," *ASTM Bulletin*, no. 214, May, 1956.

3 A. M. Freudenthal, "Safety and Structural Failure," *Trans. ASCE*, vol. 121, 1956.

4 A. M. Freudenthal, "Safety of Structures," *Proceedings, ASCE*, vol. 121, 1947.

5 J. D. North, "Some Aspects of the Relationship Between Airworthiness and Safety," *Journal, Royal Aeronautical Society*, vol. 53, 1949.

6 A. G. Pugsley, "A Philosophy of Aeroplane Strength Factors," R. M., A. R. C., London, England, no. 1906, 1942.

7 A. G. Pugsley, "Structural Safety," *Journal, Royal Aeronautical Society*, vol. 59, 1955.

8 C. R. Soderberg, "Factor of Safety and Working Stress," TRANS. ASME, vol. 52, 1930, pp. 13–28.

9 P. S. Symonds and W. Prager, "Elastic-Plastic Analysis of Structures Subjected to Loads Varying Arbitrarily Between Prescribed Limits," *Journal of Applied Mechanics*, vol. 17, TRANS. ASME, vol. 72, 1950, pp. 315–323.

10 W. Weibull, "A Survey of Statistical Effects in the Field of Material Failure," *Applied Mechanics Reviews*, vol. 5, 1952.

DISCUSSION

L. E. Iversen[1]

The paper should be a valuable contribution to the science of predicting failures inasmuch as it appears to be the first to treat the problem with more than two statistical variables based on the static failure criterion (for two variables see R. B. McCalley, Jr., "Nomogram for Section of Safety Factors," *Design News*, September 1, 1957).

Some numerical errors appear on page 389, right-hand column:

1 line 10 and 11 should read:

$$h^2 = \frac{1}{28} \times 0.1235 \times 0.1 \times 2.5^2 \times 12^2 \times 6 = 2.38,$$

$$h = 1.54 \text{ in.}$$

2 line 27 should read:

Then, with $\gamma_y = 0.20$,

3 line 32 should read:

$$h^2 = \frac{1.84}{35} \times 0.741 \times 0.08 \times 2.5^2 \times 12^2 = 2.81, h = 1.68 \text{ in.}$$

The paper gives tables for two frequency functions. For a discussion of these and others, see B. Epstein, "Statistical Aspects of Fracture Problems," *Journal of Applied Physics*, vol. 19, 1948.

A serious objection to this paper is that it treats the failure problem from a static viewpoint. This is unfortunate since almost all structures fail as a result of variable loading, variable temperature, etc., and cumulative damage. The following papers treat these problems:

I. R. Whiteman, "A Mathematical Model Depicting the Stress-Strain Diagram and Hysteresis Loop," *Journal of Applied Mechanics*, vol. 26, TRANS. ASME, vol. 81, 1959, p. 95.

A. Ferro and R. Colombo, "An Analysis of the Probability

[1] Design Analysis Section, Construction Equipment Engineering Department, International Harvester Company, Melrose Park, Ill.

Theory of Fatigue," *The Engineers' Digest*, vol. 18, November, 1957, p. 11.

R. M. Mains, "A Generalization of Cumulative Damage," TRANS. ASME, Series D, *Journal of Basic Engineering*, vol. 82, 1960, pp. 435–440.

L. Locati, "Fatigue and Other Tests With Progressively Increasing Loads," *The Engineers' Digest*, vol. 20, August, 1959, p. 8.

Ball bearing companies in general

L. G. Johnson, "Ball Bearing Engineer's Statistical Guide Book," New Departure Ball Bearings, Bristol, Conn., April, 1957.

Author's Closure

Mr. Iversen has raised several interesting points. The numerical errors, although unnecessary, did not damage the examples, which were to show the procedure of the statistical maneuver.

It should be emphasized that the physical equations expressing the relationship between R and Q in terms of their independent arguments are not necessarily "static." The adjective "physical" obviously implies a great variety of problems, and such equations can even be, say, optical or electrical in nature. It was only for the convenience of heavily occupied readers that the simple static cases were chosen as examples.

The literature mentioned by Mr. Iversen is valuable and familiar to readers in this field. There is a fundamental difference between the approach adopted by these workers and that employed in the paper. The present approach is focused on the design problem, which involves many other parameters than the pure physical ones, such as statistical (in its narrow sense), economical, and even aesthetical. The investigation of the previous workers is only concerned with the explanation of physical phenomena and the establishment of their governing equations, which only form a facet of the design problem.

As to the frequency functions, it should be noted that the proposed method is nonparametric. Although the upper and the lower bound of the probability of safety given by Tchebycheff's inequality are of great practical value, Table 1 was also intended to show the qualitative perspective of the proposed "safety" concept. Frequency functions for any particular case are of course useful. However, at the very moment, there are more important problems in design philosophy which should have priority.

B. R. SINGH
Research Engineer,
Britannia Engineering Company, Ltd.,
Titaghur, West Bengal, India

Study of Critical Velocity of Stick-Slip Sliding

The motion of a body sliding under boundary friction, at a velocity slower than a particular critical velocity of the system, proceeds in a discontinuous or stick-slip form. This paper presents a theoretical analysis and an experimental study of the critical velocity of stick-slip sliding. The influence of various parameters such as friction, damping, stiffness, and sliding mass of the system on the critical velocity was studied on a mechanical model arranged on a milling machine. The arrangement was also simulated on an electric analog computer. Good correlation was observed between the theoretical values and the experimental results on the mechanical model and on the analog computer.

RECENT INVESTIGATIONS on the sliding of a body at slow velocities under boundary friction reveal that motion is not a continuous process. Such motion proceeds with jerks. The surfaces stick together until as a result of gradually increasing pull, there is a sudden break with a consequent very rapid slip. The surfaces stick again and the process is repeated indefinitely. This intermittent motion consisting of alternate clutching and breaking away of sliding surfaces is known as "stick-slip sliding." The phenomenon is observed in many practical applications such as members sliding at low velocities in machine tools, mechanisms, and measuring apparatus. It may also occur in some other forms. Thus Bridgeman [4][1] has shown that plastic flow during shear under extremely high confining pressures was discontinuous—chattering. Again in some tensile experiments, in which steel tensile pieces were extended very slowly, it was found that the extension was not continuous but took place in a series of jumps. In many precision applications such intermittent motion, apart from causing increased wear and undesirable stresses in the sliding members, affects adversely the accuracy and sensitivity of a shifting or a movement.

In the case of machine tools, during the process of positioning a table, although the driving motor may be rotating smoothly, the motion of the driven table usually breaks down into a series of jumps and proceeds in the stick-slip form, particularly at the low speeds employed during the final adjustments. Thus in some heavy machine tools with tables weighing several tons, these jumps may extend to several millimeters. This affects positioning accuracy, resulting in dimensional error on the job, while with low feeds, it may also affect the surface finish. Then again recent developments in high-precision machine tools indicate that the automatic controls (such as E.M.I. position-control system) can provide the highest accuracies providing the machine to which they were applied is capable of utilizing the control information. It is generally recognized that the kinematic problems involved in designing such high-precision machine tools to take full advantage of these controls are by no means solved. Among the main requirements of such a machine to provide a sensitive response in the drive mechanism are the elimination of stick slip and provision of an extremely high stiffness.

Earlier studies on friction by Bowden, Leben, and Tabor [1] and [2] and Papenhuysen [3] reveal that if the sliding system possesses any degree of elastic freedom, the sliding may proceed intermittently. Such relaxation oscillations of the elastic system are possible when the kinetic friction during slip is lower than the static friction during stick. They found that, with any particular mechanical system, the type of sliding is greatly influenced by the nature of metals and the lubricant films on the surface. The nature of motion is of course also controlled by the mechanical properties of the moving system. It was observed that, if the velocity of sliding is less than or equal to a particular critical value, the motion occurred in the stick-slip form. On the other hand, if the velocity of sliding is greater than the critical, the motion occurred in the smooth form [7, 8]. Ishlinsky and Kraghelsky [10] studied the influence of the period of rest on the rise of frictional resistance during stick. Bowden, Gregory, and Tabor [11], Merchant [12], Rabinowicz [13], and others investigated the influence of certain special lubricants on the intermittent motion. Sinclair [9] reported experimental investigations on frictional vibration including stick slip. Deryaguin, Push, and Tolstoi [14] made a detailed theoretical study of the different factors that influence the critical velocity of stick slip. This paper presents an extended theoretical analysis of the latter work and

[1] Numbers in brackets designate References at end of paper.

Contributed by the Machine Design Division and presented at the Annual Meeting, Atlantic City, N. J., November 29–December 4, 1959, of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS.

NOTE: Statements and opinions advanced in papers are to be understood as individual expressions of their authors and not those of the Society. Manuscript received at ASME Headquarters, April 13, 1959. Paper No. 59—A-146.

Nomenclature

A = dimensionless parameter $= \dfrac{\nabla F}{v\sqrt{km}}$

$A_c = \dfrac{\nabla F}{v_c\sqrt{km}}$

C_1, C_2 = equivalent-viscous-damping constants

F, F_0 = kinetic and static forces of friction, respectively

k = spring scale of elastic system

$\nabla F = F_0 - F$

m = mass of sliding element

t, t_1, t_2 = time, period of slip, period of stick, respectively

v, v_c = velocity and critical velocity, respectively, of drive

$W = mg$ = weight of sliding element

x, x', x'' = instantaneous displacement, velocity, and acceleration of sliding element during sliding

$w_n = \sqrt{k/m}$ = natural circular frequency

ξ_1, ξ_2 = damping ratios

$\xi' = \xi_1 + \xi_2$

an experimental study of critical velocity. The influence of various parameters such as the difference of static and kinetic friction, damping, stiffness, and mass, on the critical velocity is studied on a mechanical model arranged on a milling machine. The arrangement is also simulated on an electric analog computer.

Critical Velocity

Consider an elastic system sliding under boundary friction over a surface which is relatively rigid. The sliding weight $W(= mg)$ is coupled by a spring of stiffness k to a drive of constant velocity v, Fig. 1. As the drive moves steadily, the motion of the mass or slip begins when the compressive force in the spring just overcomes the static frictional force F_0. During slip the frictional resistance is the kinetic friction F which has a value lower than its static value. At the end of slip the mass sticks again and the kinetic friction rises to its static value.

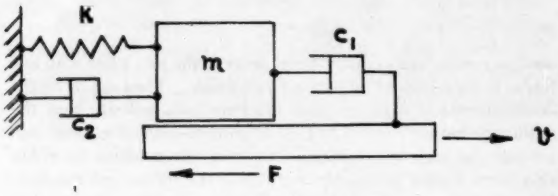


Fig. 1

The differential equation of motion of the mass during the slip may be written as

$$mx'' + (c_1 + c_2)x' + kx = \nabla F + kvt + c_1 v \qquad (1)$$

where c_1 = the coefficient of viscous damping between the sliding mass and the drive including internal material friction, structural friction, and external damper, if any; c_2 = the coefficient of viscous damping between the rubbing surfaces including external damper, if any.

Using the notation for damping ratios, i.e., the ratios of viscous damping constants to their critical values,

$$\xi_1 = \frac{c_1}{2\sqrt{mk}}; \quad \xi_2 = \frac{c_2}{2\sqrt{mk}}; \quad \xi' = \xi_1 + \xi_2$$

and initial conditions: At $t = 0$, $x = 0$, $x' = 0$, the solution of Equation (1), for the small damping ratios ($\xi_1^2 \ll 1$, and $\xi_2^2 \ll 1$) usually obtained in practice, is given by

$$x = \frac{\nabla F}{k} + vt + \frac{2\xi_1 v}{w_n} - \frac{v}{w_n} e^{-\xi' w_n t}[(A + 2\xi_1)\cos w_n t + (1 + A\xi')\sin w_n t] \qquad (2)$$

$$x' = v\{1 - e^{-\xi' w_n t}[\cos w_n t - (A + \xi' + 2\xi_1)\sin w_n t]\} \qquad (3)$$

$$x'' = \frac{\nabla F}{m} e^{-\xi' w_n t}\left[\left(1 + \frac{2\xi'}{A} + \frac{2\xi_1}{A}\right)\cos w_n t + \left(\frac{1}{A} - \xi'\right)\sin w_n t\right] \qquad (4)$$

where $\nabla F = F_0 - F$; $w_n = \sqrt{k/m}$, and A is taken as a dimensionless parameter and may be written in the form

$$A = \frac{\nabla F}{v\sqrt{km}} \qquad (5)$$

A study of the expression for the instantaneous velocity of the sliding mass for different values of the drive velocity v will reveal the following: For all values of v greater than a particular critical value, the velocity x' becomes steady after an initial transient;

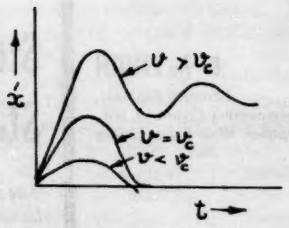


Fig. 2

on the other hand, for $v \leq v_c$, the velocity x' becomes zero and the mass sticks, thus initiating the stick slip, Fig. 2. The critical velocity may now be evaluated by noting that, in Equations (3) and (4), $x' = 0$, $x'' = 0$ for $t = t_1$, where t_1 = period of slip. This gives

$$1 - e^{-\xi' w_n t_1}[\cos w_n t_1 - (A_c + \xi' + 2\xi_1)\sin w_n t_1] = 0 \qquad (6)$$

and

$$\left[\left(1 + \frac{2\xi'}{A_c} + \frac{2\xi_1}{A_c}\right)\cos w_n t_1 + \left(\frac{1}{A_c} - \xi'\right)\sin w_n t\right] = 0 \qquad (7)$$

where the parameter $A = A_c$ under the critical condition and has been defined as

$$A_c = \frac{\nabla F}{v_c\sqrt{km}} \qquad (8)$$

where A_c can be expressed, in terms of damping ratios, from Equations (6) and (7) as

$$2(\xi_1 + \xi_2)\tan^{-1}\left[\frac{A_c + 4\xi_1 + 2\xi_2}{A_c(\xi_1 + \xi_2) - 1}\right] = \log\,[A_c^2 + 2(3\xi_1 + \xi_2)A_c + 12\xi_1^2 + 2\xi_2^2 + 10\xi_1\xi_2 + 1] \qquad (9)$$

The critical velocity for a given elastic system can now be calculated from Equations (8) and (9) knowing the values of the elastic constants of the system. Fig. 3 shows these results in a graphical form in terms of nondimensional parameter A_c and damping ratios ξ_1 and ξ_2.

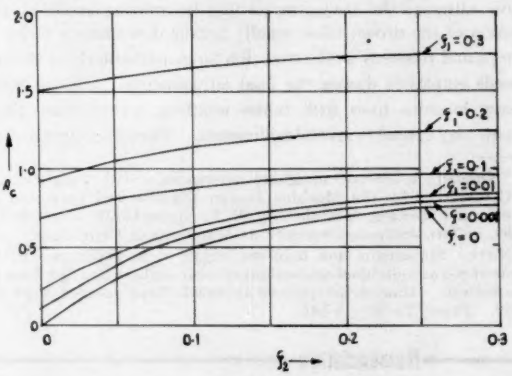


Fig. 3

In practice the critical velocity for a system should be as low as possible in order to eliminate the possibilities for the occurrence of stick slip. This may be done by reducing the difference between the static and kinetic friction, increasing damping, and the stiffness-inertia ratio.

Also the time period of slip may be obtained by solving Equation (6) for t_1. The period of stick t_2 can be obtained by noting that at the end of slip one must have

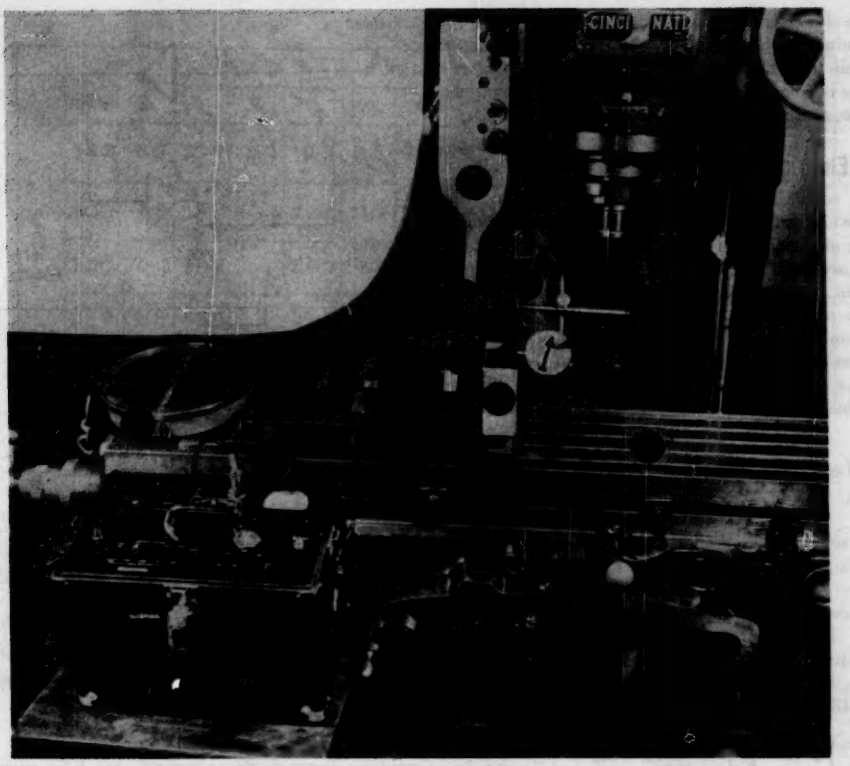


Fig. 4

1	Sliding block	5	Dial indicator	9	Milling table
2	Guide block for 1	6	Magnetic base	10	Milling head
3	Vibration pickup	7	Vibration meter	11	Spring balance
4	Oil damper	8	Clamp	12	Plate spring

$$kvt_2 = \nabla F + m\,(x'')_{t=t_1} \tag{10}$$

For a drive at critical velocity, t_2 is of course given by

$$t_2 = \frac{\nabla F}{v_c k} \tag{11}$$

Mechanical Model and Experimental Method for Critical Velocity

It was considered necessary that the experimental analysis of stick slip be conducted on a mechanical mode where the different variables affecting the process could be suitably controlled.

Fig. 4 shows a photograph of a model arranged on this basis on a milling machine. The mass m consists of a steel block weighing 5 lb, a vibration pickup (crystal type) weighing 0.5 lb, and other weights that can be added to the sliding block. The block is coupled to the milling head by a rigid attachment, through a leaf spring the stiffness of which corresponds to k. The leaf spring is clamped at its ends and loaded at mid-span. The mass is guided to slide freely in another steel block clamped to the milling table. A dashpot filled with oil is also clamped to the table. A baffle plate forming the piston for the dashpot is connected to the sliding block through a pin.

Spring scale of 133 to 600 lb/in. were obtained with different leaf springs. The damping ratios ξ_1 and ξ_2 were determined by free vibration (log-decrement) tests. For the purpose of these tests the Coulomb friction F was of course practically eliminated by sliding the mass over steel balls introduced in especially prepared V-grooves between the rubbing surfaces. The damping with the dashpot introduced and removed measures approximately $\xi' (= \xi_1 + \xi_2)$ and ξ_2, respectively. Variation in ξ_1 is obtained by changing depth of baffle plate submerged in oil. In this manner the following damping ratios were obtained

$$\begin{aligned} &\xi_2 = 0.01 \ (k = 133 \text{ lb/in., } W = 15 \text{ lb}) \\ &\xi_2 = 0.008 \ (k = 200 \text{ lb/in., } W = 15 \text{ lb}); \ \xi_1 = 0 \\ &\xi_1 = 0.1, 0.2, 0.3 \ (k = 133 \text{ lb/in., } W = 15 \text{ lb}) \end{aligned} \tag{12}$$

and $\xi_1 = 0.081, 0.162, 0.243$ $(k = 200$ lb/in., $W = 15$ lb)

The mass m participating in vibration was varied by mounting different weights to the slide block. Attempt was made in the design to apply forces through coupling the spring as near the central plane of the sliding block as possible. No additional external pressure was applied on the rubbing surfaces. The frictional forces F_0 and F are then equal, respectively, to the corresponding coefficients of friction multiplied by the weight on the rubbing surfaces. These surfaces were carefully prepared and ground fine. Paraffin oil was used as the lubricant. Under these conditions the average values of the coefficients of static and kinetic friction, obtained by measuring the deflection of leaf spring by a precision dial indicator (reading up to 1/10000 in.) are found to be 0.2 and 0.15, respectively.

The critical velocity was determined approximately by moving

the drive at the slowest velocity and increasing the velocity until stick slip is eliminated; the milling machine having a long speed range ($^5/_{16}$ in/min to 120 in/min) in small steps. The motion of the sliding mass is detected by means of the pickup, mounted on the sliding mass, and a vibration meter (G.R. type 761-A).

Simulation on Electric Analog Computer

To obtain a comparatively more accurate result on critical velocity for a wider range of damping ratios, the foregoing arrangement was also simulated on an electric analog computer. The computer used is an Electronic Differential Analyzer consisting of high-gain d-c amplifiers, attenuators, multipliers, and arbitrary function generators. The high-gain d-c amplifiers can be used as summing devices, inverters, subtracters, integrators, or differentiaters by suitable choice of input and feedback networks.

In the present problem the equation of motion of the mass is arranged as

$$x'' = -\left(\frac{c_1 + c_2}{m}\right) x' + \frac{c_2}{m} v - \frac{k}{m} (x - vt) + \frac{\nabla F}{m} \quad (13)$$

Now for any set of values of m and k, say $m = 5.5/32.2$ slugs, and $k = 4800$ lb/ft respectively, the natural circular frequency (168 radians/sec) is too fast for the servo-driven plotting device employed in these tests. This necessitated the use of a time scale factor S_t. Further, an amplitude scale factor S_v is required to keep the voltage levels in the computer within its normal operating range, i.e., 100 volts. By introducing these scale factors Equation (13) assumes the form

$$x'' = -\left(\frac{c_1 + c_2}{mS_t}\right) x' + \frac{c_2 v}{mS_t} - \frac{k}{mS_t^2} (x - vt) + \frac{\nabla F}{m} \cdot \frac{S_v}{S_t^2}$$
$$= a_1 x' + a_2 v - b (x - vt) + B \quad (14)$$

where

$$x = x/S_v; \quad a_1 = \frac{c_1 + c_2}{mS_t}; \quad a_2 = \frac{c_2}{mS_t}; \quad b = \frac{k}{mS_t^2};$$
$$\text{and} \quad B = \left(\frac{\nabla F}{m} \cdot \frac{S_v}{S_t^2}\right)$$

The simplified expression (14) has been employed for the computer setup as shown in Fig. 5. The initial conditions for the setup are—when $t = 0$, $x = 0$, and $x' = 0$. The values of computer parameters a_1, a_2, b, and B have been calculated by substituting the required quantities given in Tables 1, 2, and 3. On setting these values in the computer by means of potentiometers, the value of the velocity of drive—which corresponds to the voltage input in the computer—has been varied to obtain the necessary critical value V_c for the initiation of stick slip, see Fig. 2.

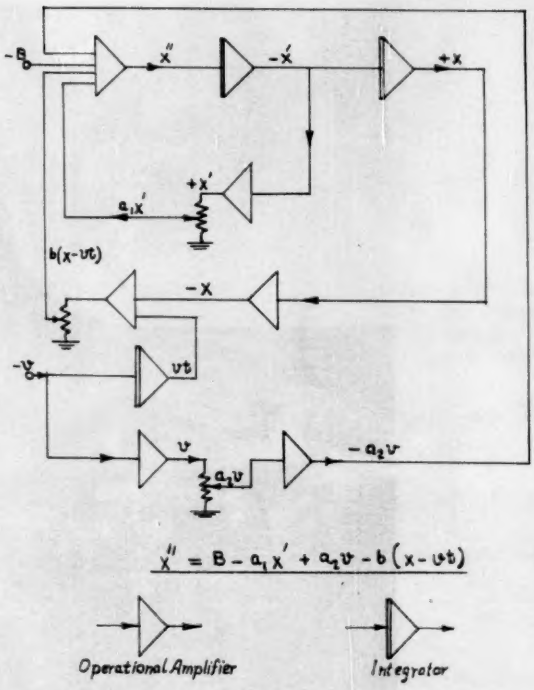


Fig. 5

Results—Critical Velocity

The critical velocities obtained for different spring scales, sliding weights, and damping ratios are represented in Tables 1, 2, and 3. The theoretical values, the values obtained on the mechanical model, and those given by the electric analog computer are indicated in these tables as V_{c_1}, V_{c_2}, and V_{c_3}, respectively. In Tables 1 and 2 the damping ratio ξ_2 remained constant at 0.01 or 0.0081 while ξ_1 was varied through 0, 0.081, 0.1, 0.162, 0.2, 0.243, and 0.3. However, with higher values of ξ_1, i.e., 0.1, 0.2, and 0.3, the verification was obtained only on the computer and the results so obtained are presented in Table 3.

Critical-velocity tests were also conducted on the model with a large number of mineral lubricants of different grades with polar

Table 1 V_c (in/min)

	K lb/in.	W lb	$\xi_1 = 0.0$			$\xi_1 = 0.1$			$\xi_1 = 0.2$			$\xi_1 = 0.3$		
			V_{c_1}	V_{c_2}	V_{c_3}	V_{c_1}	V_{c_2}	V_{c_3}	V_{c_1}	V_{c_2}	V_{c_3}	V_{c_1}	V_{c_2}	V_{c_3}
ξ_2 = 0.01	365	5.5	36.7		36.0	10.5	7.75	10.6	7.5	5.0	7.1	6.7	5.0	6.2
	267	7.5	50.0		47.5	14.2	13.0	14.2	10.2	7.7	9.2	9.3	7.8	9.0
	200	10.0	67.5		63.0	19.0	16.5	18.9	13.5	10.0	13.5	12.5	10.0	12.7
	160	12.5	83.0		78.7	23.8	20.5	24.0	17.0	13.0	17.2	15.6	13.0	16.5
	133	15.0	99.5		92.2	28.5	25.5	28.8	20.3	16.5	19.3	18.7	16.5	18.7

Table 2 V_c (in/min)

ζ_2 = 0.008	K lb/in	W lb	ζ_1 = 0.0			ζ_1 = 0.081			ζ_1 = 0.162			ζ_1 = 0.243		
			V_{c1}	V_{c2}	V_{c3}	V_{c1}	V_{c2}	V_{c3}	V_{c1}	V_{c2}	V_{c3}	V_{c1}	V_{c2}	V_{c3}
	540	5.5	20.7	16.5	19.2	8.5	7.75	9.0	6.6	5.0	5.4	5.4	5.0	5.4
	400	7.5	28.7	20.5	26.1	11.7	10.0	12.0	8.3	7.75	8.1	7.5	5.0	8.4
	300	10.0	38.2		36.0	15.6	13.0	15.0	11.2	10.2	10.8	10.0	7.75	10.2
	250	12.5	480		48.0	19.5	16.5	18.0	14.0	13.0	15.3	12.5	10.0	13.2
	200	15.0	57.5		53.0	21.7	20.5	22.5	16.8	13.0	18.0	15.0	13.0	16.2

Table 3 V_c (in/min)

ζ_2	K lb/in	W lb	ζ_1 = 0		ζ_1 = 0.1		ζ_1 = 0.2		ζ_1 = 0.3	
			V_{c1}	V_{c3}	V_{c1}	V_{c3}	V_{c1}	V_{c3}	V_{c1}	V_{c3}
0.1	300	5	0.350	0.340	0.304	0.360	0.240	0.240	0.202	0.204
0.2	300	5	0.356	0.308	0.280	0.240	0.217	0.180	0.182	0.120
0.3	300	5	0.332	0.288	0.255	0.210	0.204	0.156	0.179	0.108

and other additives in different proportions. The observations are in broad agreement with those of Bowden and Tabor [1] and Merchant [12] and are reported elsewhere. It appears that the additives lower the critical velocity by narrowing down the difference between the static and kinetic friction, the former decreasing more than the latter while damping remains practically unaltered. For instance, in the case with k = 365 lb/in., W = 5.5 lb, ξ_1 = 0.1, ξ_2 = 0.01, and carbon tetrachloride added to paraffin oil, the critical velocity decreased, approximately from 25.5 in/min to 16.4 in/min with mole fractions 0.5 and 0.75, respectively. (The decrease in the difference between F_0 and F being in proportion to the reduction in the critical velocity.)

The results of observations on critical velocity are plotted in Figs. 6 (*a*, *b*, and *c*) respectively. Fig. 6(*a*) shows the theoretical and experimental values of the critical velocity v_c (in/min) versus the spring scale k (lb/in.) for different damping ratios, Fig. 6(*b*) shows these values v_c (in/min) versus the sliding weight W (lb), and Fig. 6(*c*) the critical-velocity parameter

$$\left(\frac{v_c\sqrt{km}}{\nabla F}\right)$$

versus the damping ratio ξ_2 for different damping ratios ξ_1.

Conclusion

A theoretical and experimental study of the critical velocity of stick-slip sliding has been presented. The results for critical velocity in Figs. 6 (*a*, *b*, and *c*) show a good correlation between the theoretical values and those obtained from the experiments. They reveal the manner and extent of the dependence of the critical velocity of stick slip on the different parameters. Since stick slip occurs at a velocity lower than the critical, it is necessary in practice to have a critical velocity of the system as low as possible. This can be achieved by narrowing down the difference between the static and kinetic forces of friction, increasing stiffness-inertia ratio, and by increasing damping in the system.

The difference between static and kinetic coefficients of friction can be reduced considerably by replacing sliding friction by rolling friction or by using lubricants with special polar additives. Roller guides are successfully applied in machine tools with this object. Moreover by reducing friction in this manner, the requirement for the provision of high stiffness and power in the drive is reduced. A high stiffness in the drive may be achieved, for example, by the use of unidirectional drive which must be located toward the end of kinematic scheme, as near to the driven body as possible.

It will be seen that increase of damping in the sliding system may help to lower the critical velocity considerably. Damping can be augmented externally by the use of dampers, or by employing grease lubrication in place of fluid lubricants. On the other hand, damping may be increased by the application of special materials with high internal friction as, for instance, the use of plastics or cast iron in place of other metals.

It may also be mentioned here that a study of the influence of the fluctuations in the drive velocity or the application of a forced vibration on the sliding system reveals that resonance vibrations enhance the stick-slip effect while high-frequency vibrations may reduce the stick-slip effect [16, 17].

Acknowledgment

The author expresses his gratitude to Dr. S. R. Sengupta, Director, Indian Institute of Technology, Kharagpur, where this

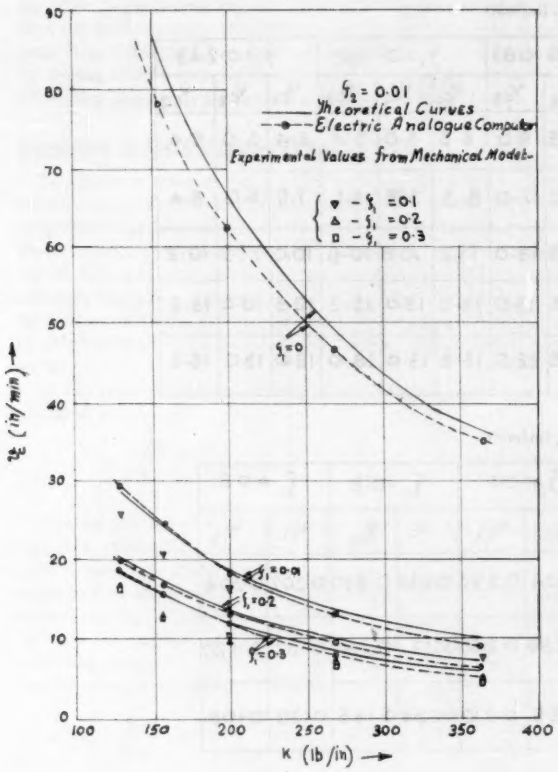


Fig. 6(a)

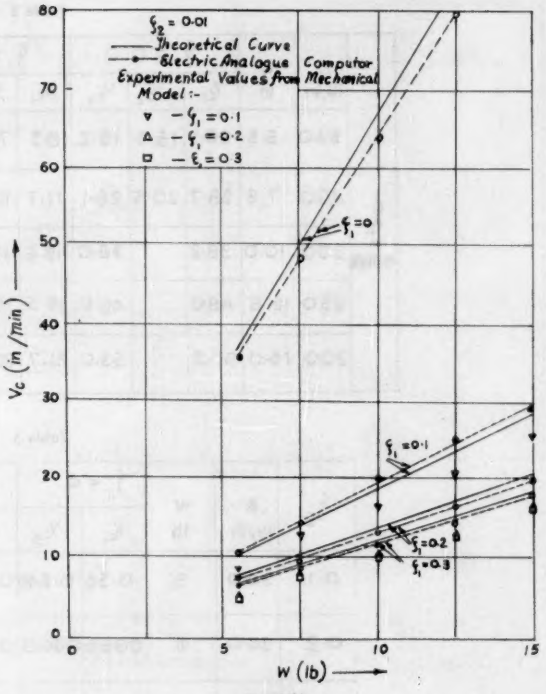


Fig. 6(b)

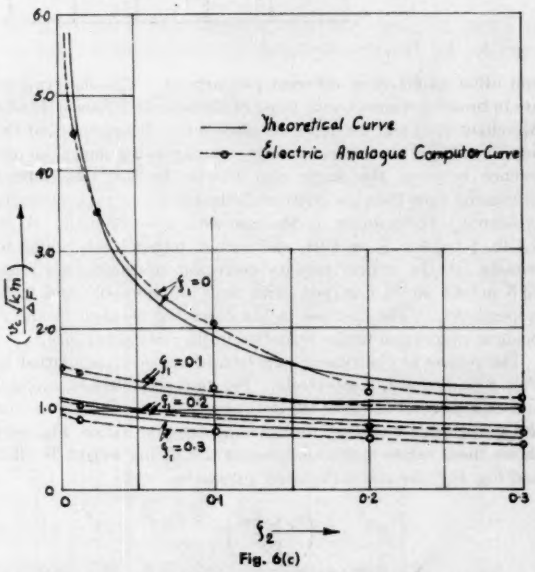


Fig. 6(c)

research was carried out, for encouragement and to Professor Dr. Tech. Sc. V. Push, of Institute of Machine Tools and Tools, Moscow, USSR (formerly at Kharagpur), for guiding this work. Thanks are due to Prof. B. M. Belgaumkar, Head of the Department of Mechanical Engineering, Indian Institute of Technology, Kharagpur, for encouragement, to the Workshop staff for providing continual assistance for mechanical work, and to Mr. H. B. Mohanti for assistance on computers.

References

1 F. P. Bowden and L. Leben, "The Nature of Sliding and the Analysis of Friction," *Proceedings of the Royal Society*, series A, vol. 169, p. 371.

2 F. P. Bowden, L. Leben, and D. Tabor, "Sliding of Metals, Frictional Fluctuations, and Vibration of Moving Parts," *Engineer* (*London*), vol. 168, 1939, pp. 214–217.

3 P. J. Papenhuysen, "Wryvingsproeven in Verband met het Slippen van Hutobanden," *De Ingenieur*, vol. 53, 1938.

4 Bridgeman, *Proceedings of the American Academy of Arts and Sciences*, vol. 53, 1918, p. 269.

5 Elam, *Zeitschrift für Physik*, vol. 88, 1934, p. 661.

6 M. Eljasberg, "About Vibration Under Boundary Friction," *Machine Tools and Tools*, vol. 15, 1951.

7 H. Blok, "Fundamental Aspects of Boundary Lubrication," *SAE Journal*, vol. 46, 1940.

8 F. Morgan, M. Muskat, and D. W. Reed, "Studies in Lubrication," *Journal of Applied Physics*, vol. 12, 1941.

9 D. Sinclair, "Frictional Vibrations," *Journal of Applied Mechanics*, vol. 22, TRANS. ASME, vol. 77, 1955, p. 133.

10 I. V. Kraghelsky, *Journal of Technical Physics* (*Moscow*), vol. 3, 1944.

11 F. P. Bowden, J. N. Gregory, and D. Tabor, *Nature*, July 28, 1945.

12 M. E. Merchant, *Lubrication Engineer*, vol. 2, no. 3, 1946.

13 E. Rabinowicz, "Friction of Boundary-Lubricated Surfaces," Second U. S. National Congress on Applied Mechanics, 1955.

14 B. Deryaguin, V. Push, and D. Tolstoi, *Journal of Technical Physics* (*Moscow*), vol. 6, 1956; Conference on Friction and Lubrication, London, England, October, 1957.

15 B. R. Singh and V. Push, "Stick-Slip Sliding," Third Congress on Theoretical and Applied Mechanics, India, 1957.

16 V. Push, B. R. Singh, and P. D. S. Varma, "Stick Slip Under Fluctuating Drive," *Journal of Science and Engineering Research*, vol. 1, pt. 2, 1957.

17 B. R. Singh and V. Push, "Stick-Slip Under Forced Vibration," *Journal of Science and Engineering Research*, vol. 2, pt. 2, 1958.

18 B. R. Singh, "Sensitivity of Slow Shifting Under Stick-Slip," *The Engineer*, London, Jnly 29, 1960.

19 B. R. Singh and H. B. Mohanti, "Experimental Investigations on Stick-Slip," *The Engineer* (*London*), April 3, 1959.

Inertial Effects in a Multiple-Ball Transmission

W. S. ROUVEROL
Associate Professor of Mechanical Engineering, University of California, Berkeley, Calif.

A new type of mechanical variable-speed transmission, which owes its improved efficiency and wear-life to the utilization of multiple point contacts in pure rolling, has a power capacity in proportion to the maximum tractive forces and rolling velocities developed by a large number of small steel balls. Designing for maximum power capacity depends on finding the point at which the components of tractive force arising from inertial effects at high speed become so large as to interfere with the proper kinematic action of the balls. Consideration is given to the influence of a number of design and operating parameters, such as ball diameter, cage diameter, cage velocities, and velocity ratio.

Introduction

DEVISING AN INFINITELY VARIABLE TRANSMISSION OR torque converter that is both compact and efficient has been a classic problem in mechanical engineering for more than a generation. To date there has been no completely satisfactory solution and consequently the field has been divided among a wide number of electrical, hydraulic, and mechanical devices, each of which appears to be the one that is least unsatisfactory for its particular use.

One of the most recent entries into this highly competitive field is the multiple-ball transmission [1].[1] This is a metallic traction device which achieves high efficiency and improved wear resistance by employing a large number of rolling point contacts between hardened steel disks and balls. The key element of this type of drive is a rotatable cage containing a large number of ordinary ball bearing balls, each of which is free to rotate in any direction in a spherical or cylindrical seat. The cage is thin enough so that the balls project slightly beyond its surface at each side. Two of the first experimental cages are shown in Figs. 1 and 2.

When such a cage is pressed between the faces of the two opposed disks, with its axis of rotation lying in the plane of the axes of the two disks, it will rotate at the average angular velocity of the two disks. Each ball moves at constant speed in a circular orbit about the center of the cage and in addition has a superimposed spinning motion that permits it to roll without slipping on the faces of both disks. The motion of the balls is in fact more truly "planetary" than that of so-called planetary gears.

The axis of rotation of the cage intersects the disk faces at the point where the velocities of the disks are equal and opposite, and if the distances from this point to the axes of the driving and driven disks are a and b, respectively, the speed ratio is a/b. By lateral shifting of the cage or of one or both disks (or of the coaxial groups of these elements in a multicage system such as shown in Fig. 3) the ratio a/b may be altered and with it the output speed.

In theory this ratio may be anything from minus infinity to plus infinity. In practice, however, although the device can produce output speeds in either direction many times the input speed, infinite output speed is naturally impossible: To achieve it would obviously require infinite power. There is, on the other hand, no physical limit to the speed reduction achievable. The output speed may be run down to zero from either the forward or

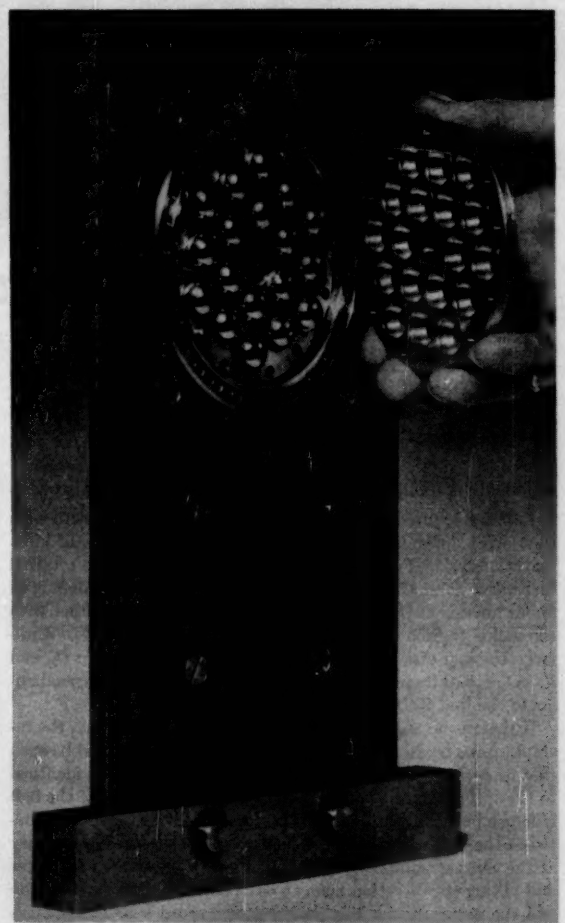


Fig. 1 Key element of the multiple-ball variable speed drive: A rotatably supported cage in which numerous balls are housed in spherical journal seats. When pressed between the faces of two disks at a point displaced from either axis, this element enables the balls to roll without sliding on both disks and to transmit power from one disk to the other with minimum frictional loss.

[1] Numbers in brackets designate References at end of paper.

Contributed by the Machine Design Division and presented at the Annual Meeting, Atlantic City, N. J., November 29–December 4, 1959, of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS.

NOTE: Statements and opinions advanced in papers are to be understood as individual expressions of their authors and not those of the Society. Manuscript received at ASME Headquarters, December 15, 1958. Paper No. 59—A-14.

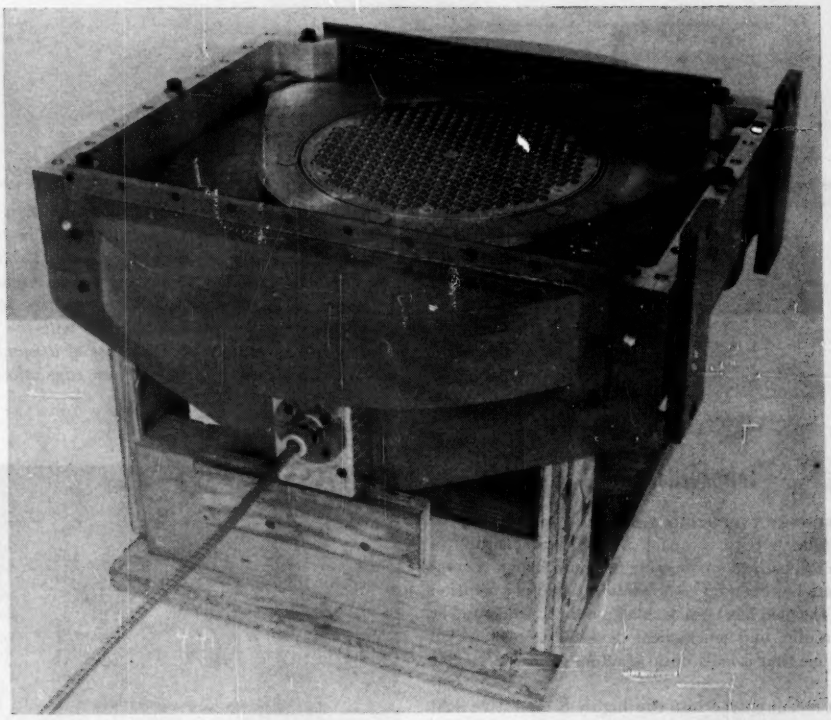


Fig. 2 Interior of a 35-hp experimental automotive transmission. The driving disk has been removed to show the 312 $^5/_{16}$-in. diam balls, which are housed in cylindrical seats in a one-piece seven-inch diameter cage.

reverse direction, in either a single cage system, or, if the disks are trunnion-mounted to avoid interferences, in a multicage system. This type of transmission is thus unique in the field of speed changers so far as range of adjustability is concerned.

Kinematic Principle

In order to understand the unique planetary motion on which the device depends, it is helpful to follow through a brief kinematic analysis. Fig. 4 shows a ball pressed between two rotating disks at a point C. The angular velocity of the front disk is ω_1 about an axis at O_1, and its velocity at the point of contact with the ball is V_A in a direction perpendicular to the line O_1C. Similarly, the rear disk has an angular velocity ω_2 about point O_2, and its velocity at its point of contact with the ball is V_B, perpendicular to the line O_2C.

If there is no slipping between the ball and either disk, the instantaneous translatory velocity and spin of the ball must be such that its surface velocity at each point of contact must be identical to the surface velocity of the disks. An enlarged view of the ball is shown in Fig. 5, with velocity vectors indicating the relation between the ball's translatory velocity V_C, and its surface velocity at the contact points, V_A and V_B. Using Cartesian co-ordinates and taking the direction from O_2 to O_1 as the positive x-direction and the perpendicular to this direction toward C as the positive y-direction, expressions for the x and y-components of V_C are as follows:

$$V_{Cx} = \frac{1}{2}(V_{Ax} + V_{Bx}) \tag{1}$$

$$V_{Cy} = \frac{1}{2}(V_{Ay} + V_{By}) \tag{2}$$

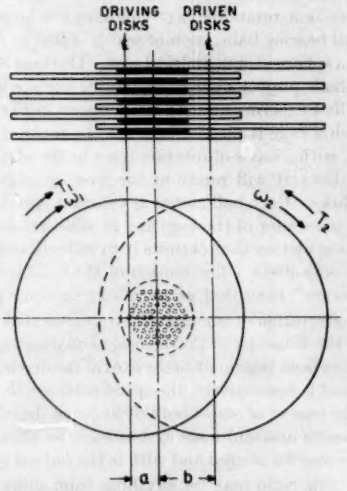


Fig. 3 Plan and elevation of the disks and cages of a multicage transmission. (Rotatable support not shown.)

The vector sum of these components gives the magnitude of V_C:

$$V_C = \frac{1}{2}\sqrt{(V_{Ax} + V_{Bx})^2 + (V_{Ay} + V_{By})^2} \tag{3}$$

This translatory velocity makes an angle ϕ with the positive x-direction such that:

$$\tan\phi = \frac{V_{Ay} + V_{By}}{V_{Ax} + V_{Bx}} \tag{4}$$

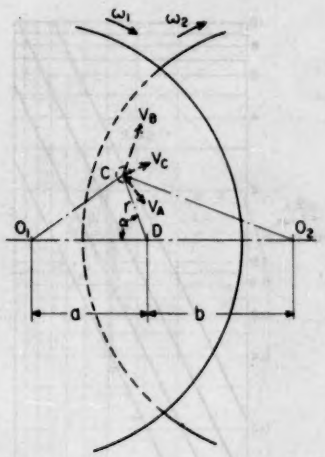


Fig. 4 Diagram showing the translatory velocity of a ball pressed between two rotating disks

Returning to Fig. 4, if the point C is connected to some point D on the line O_1O_2 a distance a from O_1 and b from O_2, the length of the line from C to D designated as r, and the angle it makes with the line O_1O_2 as α, then the x and y-components of the common velocities at the contact points may be expressed as follows:

$$V_{A_x} = -\omega_1 r \sin \alpha \tag{5}$$

$$V_{B_x} = -\omega_2 r \sin \alpha \tag{6}$$

$$V_{A_y} = \omega_1(-a + r \cos \alpha) \tag{7}$$

$$V_{B_y} = \omega_2(-b + r \cos \alpha) \tag{8}$$

Adding equations (5) and (6) gives:

$$V_{A_x} + V_{B_x} = -(\omega_1 + \omega_2) r \sin \alpha \tag{9}$$

If the point D has been chosen so as to divide the line O_1O_2 in inverse proportion to the angular velocities ω_1 and ω_2, such that ω_2/ω_1 equals $-a/b$ (where b is considered to extend in the negative direction from the origin at D), then

$$\omega_1 a = -\omega_2 b \tag{10}$$

In this case the sum of Equations (7) and (8) becomes:

$$V_{A_y} + V_{B_y} = (\omega_1 + \omega_2)\, r \cos \alpha \tag{11}$$

Since $\sin^2 \alpha + \cos^2 \alpha = 1$, the squaring of Equations (9) and (11) and substituting into Equation (3) gives the results:

$$V_C = r\left(\frac{\omega_1 + \omega_2}{2}\right) \tag{12}$$

This is the critical relation on which the functioning of the transmission depends. It shows that the velocity of the ball is independent of the angle α which defines its angular position relative to the point D. Further, if the distance r is held constant, the velocity of any and all balls at this distance from point D must be the same. The fact that the distance r may be held constant without inducing sliding is demonstrated by dividing Equation (11) by Equation (9), which gives the result:

$$\text{ctn } \alpha = -\frac{V_{A_y} + V_{B_y}}{V_{A_x} + V_{B_x}} \tag{13}$$

Comparison with Equation (4) shows that $\tan \phi = -\text{ctn } \alpha$, and hence the instantaneous velocity of the ball V_C must be perpendicular to the line CD. It follows that the ball may be constrained to roll in a circular orbit about point D without sliding on either disk [2].

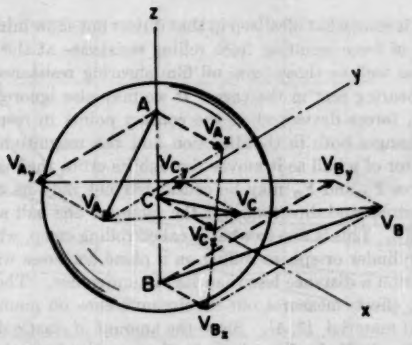


Fig. 5 Enlarged isometric diagram of a ball rolling between two moving planes, showing the relation between translatory velocity (V_C) and surface velocity at opposite points (V_A, V_B)

Tractive Force

How the device transmits torque is perhaps even less apparent than how the balls are able to move in pure rolling. Analysis proceeds from the fact that motion and forces in one plane are entirely independent of motion and forces perpendicular to that plane, from which it follows that the circular translation of the balls has no direct relation to the transmission of torque. Centrifugal forces resulting from the circular motion are quite incidental, tending merely to impose additional components of load on the spherical or cylindrical journal seats of the cage. So far as the transmission of power is concerned, the important forces on the balls lie in a plane normal to the cage; these forces are shown in Fig. 6.

Pressing of the disks against the cage produces equal and opposite normal forces N_A and N_B at the contact points. If in addition torques are applied to the driving and driven disks, equal tractive forces F_A and F_B will be developed at the contact points. The reaction to these two forces, $2F$, is exerted by the cage. As the ball spins through its orbit, these forces do not vary appreciably in either magnitude or direction, but lie always in a plane normal to the line of centers of the disks. It may be shown that due to symmetry about this line of centers, any components of force not lying in this plane would produce a slight change in velocity of the cage in such a direction as to reduce them rapidly to zero.

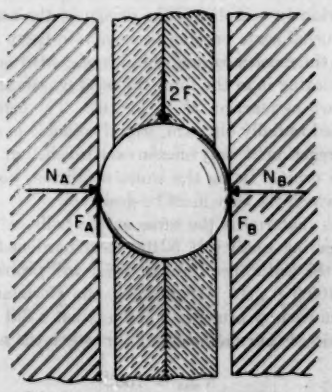


Fig. 6 Diagram of the forces acting on a ball pressed between two plates and constrained to rotate within a spherical journal seat in a cage

Fig. 6 is somewhat idealized in that it does not show minor components of force resulting from rolling resistance at the contact points, as well as those from oil film shearing resistance in the sliding bearing seat in the cage. If we may also ignore, for the moment, forces developed at the contact points in response to cyclic changes both in the direction and the magnitude of the spin vector of a ball as it moves through its orbit, the main tractive forces F_A and F_B may be considered not only as constant in magnitude and direction, but the same for one ball as for all other balls. This is due to what is called rolling creep, whereby a loaded cylinder or sphere rolling on a plane traverses with each full rotation a distance less than its circumference. The rolling body, in effect, measures out its circumference on momentarily stretched material [3, 4]. Since the amount of elastic deformation under the body depends on both normal and tractive force, the rate of creep will also depend on both. In addition, there may be at high loads a small amount of shearing action within the thin oil film in the Hertzian contact area

Rolling creep, from whatever cause, is a most convenient phenomenon for this type of transmission, since it results in a tendency for all the balls in a cage, or all the cages in a multicage system, to share the tractive load more or less equally. Lightly loaded balls tend to creep less and hence to push their portion of the driven disk ahead at a faster rate, which in turn tends to relieve the other balls of enough load to equalize the creep rates. Even though the creep rate tends to be small for steel balls rolling on steel plates (a fraction of one per cent), the spin velocity of the balls is ordinarily so high that inequalities in tractive forces are corrected almost instantaneously.

It is also of interest to note that power is not transmitted directly through the balls from one disk to another. Part of it is transferred in this way, but part of it is transmitted through the cage. If the force system of Fig. 6 is allowed to translate in the direction of the tractive friction forces, it will be evident that the force $2F$, representing the pressure of the ball against its cage seat, is doing work on the cage. But for each ball with a component of translation in the direction of its tractive forces, there is a ball on the opposite side of the cage with an equal component directed opposite to the tractive forces, so that the work being done on the cage by the balls in one half of it is being delivered back to other balls in the opposite half of the cage, and from these balls to the driven disk.

Centrifugal Reactions

Since all centrifugal forces exerted by the balls at high speed are carried by the cage rather than the disks, no loss of torque capacity is associated with this particular type of acceleration. Such forces can become quite large, however, and may seriously complicate the design of the cage if force-fed spherical journal seats become necessary to carry the load.

Probably the most meaningful index of centrifugal effect is obtained by relating it to the force imposed by the transmission of torque (shown as reaction $2F$ in Fig. 6). The centrifugal force, being directed radially outward, would coincide in direction with the bearing reaction $2F$ only once in each revolution, at the bottom of the ball's orbit. Being the worst condition, however, this is the one for which the cage must be designed.

If the disk material is the same as the balls, which will ordinarily be bearing quality E 52100 steel hardened to Rockwell "C" 62-65, the endurance limit in shear will probably be in the vicinity of 100,000 psi. According to the Hertzian analysis for the stress induced by pressing a ball against a flat surface [5], a shearing stress of this magnitude occurs when the normal force is

$$N_{max} = 160d^2 \tag{14}$$

where the normal force N is in pounds and the ball diameter d is in inches. Since this expression is based on a fatigue strength, it represents a maximum value for continuing operation, but one which may be exceeded by occasional shock or starting loads.

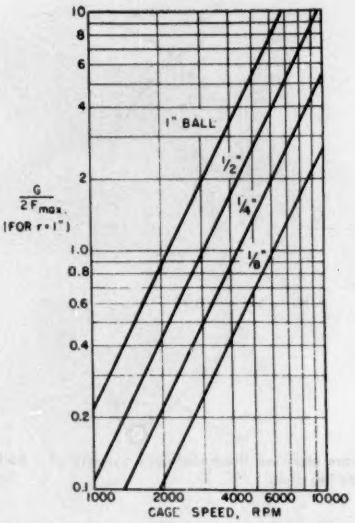


Fig. 7 Plot of Equation (15) showing ratio of centrifugal force to useful bearing load versus cage speed for various ball sizes, with ball orbit radius of 1 inch

Most of the transmissions of the multiple-ball type designed or operated to date impose normal loads N of 10 to 20 times the tractive force F. Using a fairly conservative normal load factor of 16, and computing the centrifugal force G in pounds on the basis of an assumed density for steel balls of 0.286 pounds per cubic inch, we find the ratio of the centrifugal force to maximum useful bearing load $2F$ is as follows:

$$\frac{G}{2F_{max}} = 0.213 \times 10^{-6} r d N_c^2 \tag{15}$$

where r is, as before, the radius of the ball orbit in inches, and N_c is the cage speed in rpm. This expression is plotted in Fig. 7, to show the circumstances under which centrifugal force may become a problem.

An analysis of other types of inertial effects indicates that cage velocities of 3000 to 5000 rpm are practical, as well as useful for optimizing power capacity and efficiency. If even at these speeds the $G/2F$ ratio can be kept low, preferably in the vicinity of 0.5 or less, it becomes possible to use a one-piece cage in which the balls have cylindrical seats. Because of the narrowness of their oil wedge, such seats are exceptionally efficient at light loads and high speeds; furthermore, they permit the use of a type of cage costing only a fraction of the cost of the two-piece spherical seat cage shown in Fig. 1.

It should be noted that Fig. 7 is plotted for an assumed ball orbit radius of one inch. Since this variable appears in the first power in Equation (15), Fig. 7 may be conveniently used to determine permissible $G/2F$ ratios for higher values of r by simple proportion. From the curves it is evident that for relatively small cages of about 2 inches in diameter, ball sizes of much larger than $1/4$ in. may begin to introduce centrifugal problems at high speed; likewise with cages 4 in. in diameter, the optimum ball size is evidently $1/8$ in. or less.

Tractive Forces

There is no way to entirely prevent the loss of a certain amount

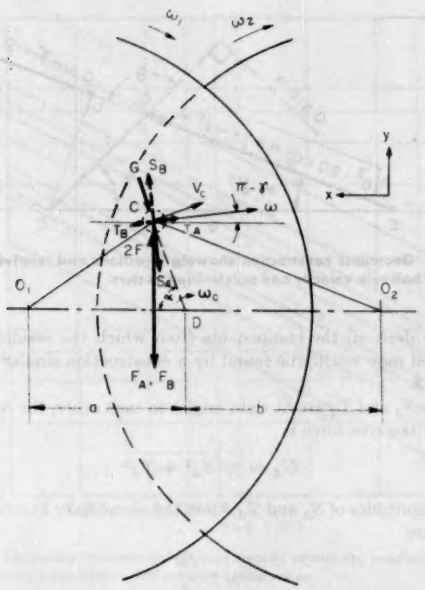


Fig. 8 Forces acting on the ball of Fig. 4: Useful tractive forces (F_A, F_B), centrifugal reaction (G), acceleration reactions (S_A, S_B), and gyroscopic reactions (T_A, T_B). Translatory velocity (V_C), spin vector (ω), and precession angle (γ) are also indicated.

of tractive capacity through inertial effects. Since the spin of the balls varies both in direction and magnitude as they move around their orbits, forces must be developed at the traction points to produce these variations. These forces do no useful work, but simply constitute undesirable components which add vectorially to the useful tractive forces F_A and F_B.

The orientation of these tractive forces is shown in Fig. 8. In addition to F_A and F_B, equal and opposite tractive forces S_A and S_B separated by the ball diameter d comprise a couple inducing spin acceleration $\dot{\omega}$; this couple acts in the plane normal to the direction of the spin vector ω. In a plane normal to this one, and hence containing the spin vector, is another pair of equal and opposite forces T_A and T_B; these forces are also separated by the ball diameter and comprise a second couple tending to produce a gyroscopic precession of the spin vector or, to be more exact, to be developed as a reaction to a forced precession of this vector.

In investigating the tractive forces resulting from ball inertia, forces due solely to acceleration and those due to gyroscopic effect may be treated independently and the results added by superposition. This is possible because the product of inertia term which arises in the more general treatment vanishes in the case of a symmetrical body such as a homogeneous sphere. The sign conventions will be the same as before: Upward and to the left for linear displacements and velocities, clockwise for angular displacements, velocities, and accelerations, are positive. A right-hand rule will be employed with rotational vectors.

Ball Spin Velocities

Returning to Figs. 4 and 5, the x and y-components of ball spin velocity ω in radians per second are as follows:

$$\omega_x = \frac{1}{d}\,(V_{A_y} - V_{B_y}) \tag{16}$$

$$\omega_y = \frac{-1}{d}\,(V_{A_x} - V_{B_x}) \tag{17}$$

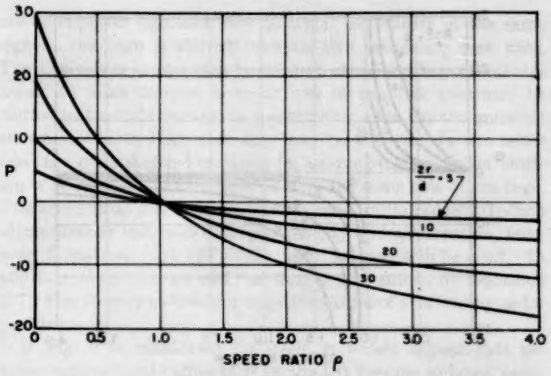


Fig. 9 P versus speed ratio for various ratios of cage diameter to ball diameter

Using again Equations (5) through (8), we find

$$\omega_x = \frac{1}{d}\,[(\omega_1 - \omega_2)r\cos\alpha - \omega_1 a + \omega_2 b] \tag{18}$$

$$\omega_y = \frac{1}{d}\,(\omega_1 - \omega_2)r\sin\alpha \tag{19}$$

Noting from Equation (12) that the angular velocity of the cage ω_c is the average of ω_1 and ω_2, and substituting the velocity ratio ρ for ω_2/ω_1, the following relations may be obtained:

$$a = \frac{\rho(a - b)}{1 + \rho} \tag{20}$$

$$\omega_1 = \frac{2\omega_c}{1 + \rho} \tag{21}$$

$$\omega_2 = \frac{2\omega_c\rho}{1 + \rho} \tag{22}$$

$$\omega_1 - \omega_2 = \frac{2\omega_c(1 - \rho)}{1 + \rho} \tag{23}$$

Using Equation (10) and the equations just given, we may rewrite Equations (18) and (19) as follows:

$$\omega_x = \frac{2\omega_c r(1 - \rho)}{d(1 + \rho)}\left[\cos\alpha - \frac{2\rho(a - b)}{(1 - \rho^2)r}\right] \tag{24}$$

$$\omega_y = \frac{2\omega_c r(1 - \rho)}{d(1 + \rho)}\sin\alpha \tag{25}$$

The following dimensionless constants may be substituted to simplify the calculations:

$$P = \frac{2r(1 - \rho)}{d(1 + \rho)} \tag{26}$$

$$Q = \frac{2\rho(a - b)}{(1 - \rho^2)r} \tag{27}$$

Using these equations, the magnitude of the ball spin becomes:

$$\omega = \omega_c P\sqrt{Q^2 - 2Q\cos\alpha + 1} \tag{28}$$

The angle made by this spin vector with the positive x-direction (see Fig. 8) is:

$$\gamma = \tan^{-1}\frac{\omega_y}{\omega_x} = \tan^{-1}\left(\frac{\sin\alpha}{\cos\alpha - Q}\right) \tag{29}$$

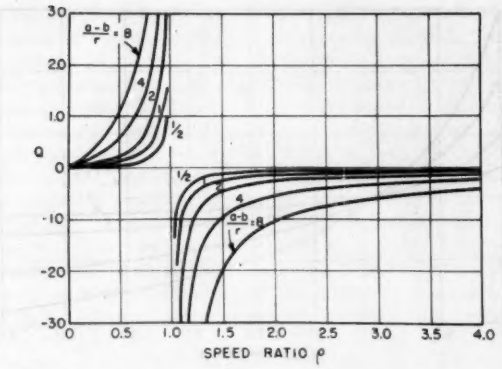


Fig. 10 Q versus speed ratio for various ratios of shaft offset to cage radius

Equations (28) and (29) reveal an interesting feature of the device, namely, that the spin axes for all balls always pass through a common point on the line of centers of the disks. Fig. 11 has been prepared in dimensionless form to show how the arguments of Equations (28) and (29) vary with α. If all sides of the triangle in this figure are multiplied by r, the illustration becomes a physical representation of the motion of a ball at radius r from the center of the cage. It will be evident from this figure and from Equations (28) and (29) that the distance to the common point E will then be $Q \cdot r$, and since the expression for Q, given in Equation (27), has r in the denominator, the distance from the center of the cage to the point E will be independent of r, and dependent only on the speed ratio and on the offset between shaft axes, which in most designs will be a constant.

Since the side EC of the triangles in Fig. 11 not only indicates the direction of the ball spin vector, but the relative magnitude as well, it is a convenient aid for visualizing the motion of the balls. The point E coincides with D when the speed ratio is zero, moves to the left as the speed ratio increases, goes to infinity in the positive x-direction at unity ratio, then moves in from negative infinity at the right as the ratio increases to values above unity. This effect is also evident in Fig. 10. It should of course be noted, however, that, since the actual spin velocity according to Equation (28) varies with P as well as Q, the scale to which the distance EC represents the spin velocity alters with the speed ratio. Since Q goes to infinity when ρ equals unity, and the scale factor, which is proportional to P, goes to zero at the same time, the true vector length becomes indeterminate. For this particular case ω can be obtained by putting P under the radical and substituting the values of Equations (26) and (27). This leads to a result which shows that at unity velocity ratio the ball spin velocity is independent of α, remains parallel to itself always in the x-direction, and has a constant magnitude equal to:

$$\omega = \omega_c \cdot \frac{a - b}{d} \tag{30}$$

At this ratio there are consequently no tractive forces induced by ball inertia. This is also true at a speed ratio of minus one, in which case ω_c goes to zero and the ball spin velocity is easily computed from the geometry of the cage.

Resultant Inertial Tractive Force

Since S_A and T_A both act at the contact point A between the ball and the driving disk, the clearest conception of their combined effect may be had by resolving them and comparing their resultant to the tractive force F_A, which acts at the same point. Since the angle between F_A and the resultant of S_A and T_A will

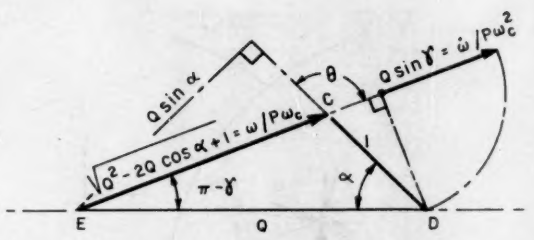


Fig. 11 Geometric construction showing directions and relative magnitudes of ball spin velocity and acceleration vectors

also be derived, the components from which the resultant was obtained may readily be found by a construction similar to that of Fig. 8.

Since S_A and T_A are at right angles to each other, the resultant inertial tractive force is

$$U_A = \sqrt{S_A^2 + T_A^2} \tag{31}$$

The magnitudes of S_A and T_A, from the elementary kinetic equations, are

$$S_A = \frac{I\dot{\omega}}{d} \tag{32}$$

and

$$T_A = \frac{I\omega\dot{\gamma}}{d} \tag{33}$$

where I is the mass moment of inertia for a solid homogeneous sphere about a diameter.

Differentiation of Equation (28) with respect to time gives the angular acceleration of the ball:

$$\dot{\omega} = \frac{\omega_c^2 PQ \sin \alpha}{\sqrt{Q^2 - 2Q \cos \alpha + 1}} = \omega_c^2 PQ \sin \gamma \tag{34}$$

Differentiation of Equation (29) with respect to time gives the precession rate:

$$\dot{\gamma} = \frac{\omega_c(1 - Q \cos \alpha)}{Q^2 - 2Q \cos \alpha + 1} \tag{35}$$

Substitution of Equations (34) into (32), (28) and (35) into (33), and (32) and (33) into (31), gives the resultant inertial tractive force

$$U_A = \frac{I\omega_c^2 P}{d} = 0.852 \times 10^{-6} r d^3 N_c^2 \left(\frac{1 - \rho}{1 + \rho}\right) \tag{36}$$

which makes an angle with the spin vector

$$\theta = \tan^{-1} \frac{S_A}{T_A} = \tan^{-1} \left(\frac{Q \sin \alpha}{1 - Q \cos \alpha}\right) = \gamma - \alpha \tag{37}$$

This is an astonishingly simple result, considering the complexities of the ball motion and the amount of manipulation involved in the derivation. Not only is U_A independent of the position of the ball in the orbit, but it remains always in the plane normal to the disks containing the ball orbit radius r. It is directed toward the cage axis D when the speed ratio is less than unity, and in the opposite direction when the ratio is greater than unity. The corresponding resultant tractive force U_B at the contact point B between the ball and the driven disk is equal in magnitude and opposite in direction to U_A, and with it comprises a couple with the ball diameter d as its arm.

As before, it is most meaningful to relate this inertial effect to available tractive force. This may be done by dividing Equation

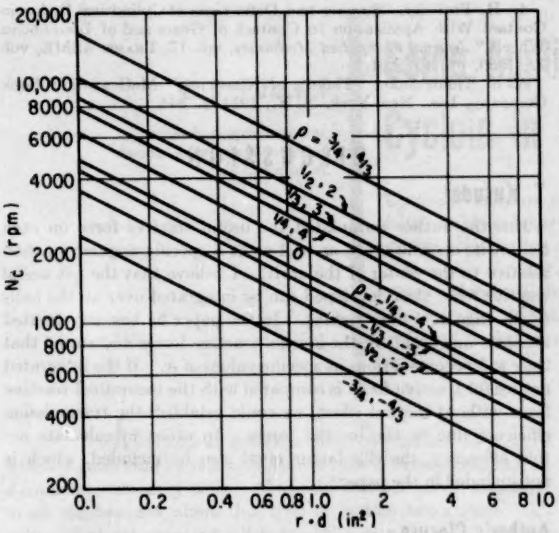


Fig. 12 Maximum recommended cage speeds versus the product of ball diameter and cage radius, for various speed ratios

(36) by the equation for allowable tractive force on which the plot of Equation (15) for centrifugal effects was based, namely

$$F_{max} = Kd^2 \tag{38}$$

where K for steel balls and disks as has been noted is about 10 and has the units pound per square inch. The resulting equation is

$$\frac{U_A}{F_{max}} = \frac{I\omega_c^2 r(1-\rho)}{5d^4(1+\rho)} = 0.852 \times 10^{-7} rdN_c^2 \left(\frac{1-\rho}{1+\rho}\right) \tag{39}$$

The most serious condition will be when the force U_A is in the same direction as F_A, as occurs when $\alpha = \pi/2$, and when U_B is in the same direction as F_B, at $\alpha = 3\pi/2$. In either of these positions, only the algebraic difference between F_{max} and U_A will be available for transmitting torque. Since Equation (36) shows that U_A increases as the square of cage velocity, this difference diminishes as cage speed increases; when it reaches zero, traction is lost at the top and bottom of the cage simultaneously and slipping for all the balls may then ensue.

Thus it is evident that at some very high speed the torque capacity of a unit, and hence also its power capacity, reduces to zero. Similarly, when the velocity of the disks is zero, the power capacity is again zero. In between, however, a maximum must be reached. By differentiation and setting equal to zero the product of cage velocity N_c and the difference between F_{max} and U_A, which is linearly related to the power capacity, a good approximation of the cage speed at maximum power may be obtained. It is found to occur when U_A reaches a value of one third of F_{max} and at any particular speed ratio depends only on one parameter, r times d:

$$N_c = 1980 \sqrt{\frac{1}{rd}\left|\frac{1+\rho}{1-\rho}\right|}, \quad \rho \neq \pm 1, \tag{40}$$

This equation is plotted as Fig. 12.

It should perhaps be noted that all of these equations are general enough to apply regardless of how a and b are varied. The most common construction is probably the one in which a and b are of opposite sign and vary in equal increments of the same sign as the cage is shifted between two stationary disk axes. The analysis is in this case based on a set of shifting co-ordinate axes. In some designs, however, one of the disk axes may be shifted instead of the cage, or occasionally, as in the transmission shown in Fig. 2, both disk axes may be shifted. In the latter case, the greatest speed variation for a given amount of axial shifting is afforded by having both axes on the same side of the cage. The speed ratio ρ is thus negative in most of its range. To find instantaneous values of ball spin velocity and acceleration, however, Equations (26), (27), (28), and (34) may still be used. In addition, since Q is an odd function of ρ as shown by Equation (27), Fig. 10 may be used for negative values of ρ as readily as for positive ones.

If Fig. 9 is considered by itself, it would appear that because extrapolated values of P on the left become so large, negative speed ratios would give very high values of inertial tractive force according to Equation (36). This conclusion might also be drawn from Fig. 12. This is not however the case because, when ω_2 is negative, the cage velocity, as based on Equation (12), becomes small. A clearer conception of the permissible operating speeds may in this case be obtained by rewriting Equation (36) in terms of the input speed ω_1. This may be done with the aid of Equations (21) and (26) to give a result showing that the force U_A is independent of the rotational direction of the output:

$$U_A = \frac{I\omega_1^2 r(1-\rho^2)}{2d^2} = 0.213 \times 10^{-6} rd^3 N_1^2 (1-\rho^2) \tag{41}$$

From this an equation analogous to Equation (40) may be obtained, to show the maximum recommended driving disk speed, in rpm, for output rotations in either direction:

$$N_1 = \frac{3960}{\sqrt{rd|1-\rho^2|}}, \quad \rho \neq \pm 1, \tag{42}$$

This equation could be plotted to give a family of curves as in Fig. 12, but in view of the ease of converting N_1 to N_c by means of Equation (21), these additional curves would essentially duplicate the information of Fig. 12.

Conclusions

Designing a multiple-ball transmission to carry a particular peak power load is not essentially different, so far as approach is concerned, from designing any other type of friction transmission. Since power capacity is a function of the product of tractive force and rolling velocity, the greatest economy in weight, size, and cost will generally be achieved by reducing the tractive force and increasing the velocity to the limits of practicability. The analysis shows that upper limits on velocity are imposed by inertial effects acting on the balls, and indicates further how these effects may be minimized so as to permit operation at the highest possible speeds.

One important variable is ball size. The number of uniformly sized balls which may be located in a given projected area of cage is inversely proportional to the square of the ball diameter. Further, the allowable normal load on each ball varies directly as the square of the ball diameter, as is evident from Equation (14). Hence the maximum total normal load on a given projected area of cage is independent of ball diameter. If the coefficient of tractive friction may be assumed to be also independent of ball size, the total tractive force which may be developed per unit projected area of cage is not affected by whether the balls are large or small, provided the speed is slow.

The analysis, however, shows that as the speed is increased, ball spin acceleration and gyroscopically induced couples produce significant nonuseful tractive forces at the contact points with the

disks. The ratio of these nonuseful tractive forces to the useful ones at a particular cage velocity and speed ratio increases directly as the ball diameter. Centrifugal forces exerted by the balls against their cage seats also increase directly with ball diameter. Equations (39) and (15), respectively, express these effects quantitatively and emphasize the importance of using a large number of small balls rather than vice versa.

A similar situation exists with regard to the radius of the cage. The same equations reveal that the balls should be clustered as closely as possible to the center of the cage, if maximum speeds are to be achieved. This fact also influences the design, in that it indicates that the cage should always be mounted peripherally as shown in Figs. 1 and 2. Although it is kinematically possible and in some ways mechanically attractive to mount the cage on a central shaft, which may be passed through a large enough hole in one of the disks to allow a reasonable amount of transverse shifting of the cage axis relative to the disk axes, such an arrangement results in a most unfavorably large radius for even the innermost ball orbits.

Fig. 12 summarizes the most important relations between the variables which affect the allowable operating speeds. It shows, for example, that, if balls of $^1/_8$ to $^1/_4$-in. diameter are used in a cage of small radius, cage velocities may easily run into the range between 3000 and 5000 rpm, depending on the limiting speed ratios to be reached at these cage speeds. In this connection it may be observed that if these cage speeds are reached when the device is idling with no load on the output shaft, in any design in which there is a torque-proportional normal loading system, then a minimum normal loading of about one tenth the maximum loading must always be applied to the cage to prevent scuffing of the disks through excessive slip arising from ball inertia.

References

1 W. S. Rouverol, "A New Infinitely-Variable Transmission," *Engineering*, vol. 183, no. 4762, June 14, 1957, p. 754.

2 After the foregoing analysis was completed it came to the attention of the writer that a mathematical derivation encompassing the relations up to but not beyond Equation (13) was first accomplished in 1933 by Dr. Hans Schnabel of Germany. No publication was ever made, however.

3 O. Reynolds, "On Rolling Friction," *Philosophical Transactions of the Royal Society of London*, vol. 166, 1876, pp. 155–174.

4 H. Poritsky, "Stresses and Deflections of Cylindrical Bodies in Contact With Application to Contact of Gears and of Locomotive Wheels," *Journal of Applied Mechanics*, vol. 17, TRANS. ASME, vol. 72, 1950, pp. 191–201.

5 S. Timoshenko, "Theory of Elasticity," McGraw-Hill Book Company, Inc., New York, N. Y., 1934, p. 344.

DISCUSSION

T. Kusuda[2]

Has the author formulated the useful tractive force on each ball at its specific orbit, as well as at a specific angular location relative to the center of the orbit? I believe that the net useful tractive force thus developed can be integrated over all the balls in the cage at a time instant. In the paper he has only treated absolute magnitude of the inertial tractive forces and stated that they will become serious at specific values of α. If the integrated net useful tractive force is compared with the theoretical tractive force without inertial effect, we could establish the transmission efficiency due to the inertial forces. In order to calculate actual efficiency, the slip factor must also be included, which is not included in the paper.

Author's Closure

I wish to thank Mr. Kusuda for his discussion and to make the following notes regarding his comments:

(1) There is no rigorous way to make a direct integration of the useful tractive forces, since the force system is redundant to the degree $(n - 1)$, where n is the total number of balls. Some approximate values can be gained by considering rolling creep, lateral forces due to misalignment, and centrifugal effects, but it is probably more practical to seek empirical results.

(2) Inertial tractive forces are independent of α, as shown in Equation (36).

(3) Inertial effects have no influence on efficiency.

(4) The slip factor varies from one part in 10,000 to one part in 1000, depending on the load, so it is a negligible part of the losses.

[2] Staff Engineer, Research and Development Section, Worthington Corporation, Harrison, N. J.

E. P. POLLITT

Senior Research Engineer,
Armour Research Foundation,
Chicago, Ill.

Some Applications of the Cycloid in Machine Design

Cycloidal motions have many applications in mechanical engineering. The multilobed epicycloid has sharply pointed cusps; therefore, a machine element performing an epicyclic motion can be utilized for performing operations requiring a corresponding action, like folding of flexible materials or feeding of components from a stack. A speed reducer, distinguished by extreme compactness, is based on the use of an epitrochoidal cam. Linkages which perform approximately straight-line motions can be designed so that they approximate the motion of the center of the generating circle of an orthocycloid.

A CYCLOIDAL CURVE is generally defined as a curve which is generated by points of a circle rolling without slipping on a coplanar stationary circle. Cycloidal motions occur, therefore, in all mechanisms where this type of motion takes place. All points of planet gears, or rollers in antifriction bearings, perform cycloidal motions. The multilobed epicycloid has sharply pointed cusps; a motion having this path is desirable in the operation of folding of flexible materials like paper or cloth, has the advantage that it can be produced by uniformly moving elements, and it is used extensively in newspapers folders. In stack feeders this motion is desirable for suction cups which pick up the individual articles from the stack. An epitrochoidal cam constrained to perform a wobble motion, used in conjunction with a stationary mangle gear, constitutes a highly efficient, compact speed reducer. Other examples of the application of cycloidal motion are function generators, indexing devices, pumps, straight-line linkages, etc.

Properties of Cycloidal Curves

The equations of cycloidal curves may be derived from the relation that the length of arc of the base circle on which the generating circle rolls is equal to the length of the arc traveled by a point on its circumference. Expressed by the angle of rotation θ of the center of the generating circle as parameter, the co-ordinates of these curves are

$$x = (R \pm r)\cos\theta \mp r\cos\frac{R \pm r}{r}\theta \qquad (1)$$

$$y = (R \pm r)\cos\theta - r\sin\frac{R \pm r}{r}\theta \qquad (2)$$

where R is the radius of the base circle, r the radius of the rolling circle. The upper signs apply to the epicycloid generated by the rolling circle rolling on the outside of the base circle, Fig. 1(a), the lower signs to the hypocycloid generated by the rolling circle rolling on the inside of the base circle, Fig. 1(b). The slope

$$\psi = \theta \pm \frac{R}{2r}\theta$$

At the cusp $[\theta = (R/r)2\pi]$ the slope becomes equal to the angle $(\theta \pm \pi)$. For $R = r$, the epicycloid has a single cusp and is called cardioid which is also one of Pascal's curves, Fig. 1(c).

For $r = \infty$ the epicycloid becomes the involute of the base circle.

For $R = \infty$ the epicycloid becomes the orthogonal cycloid.

For $r = R/2$, the equation of the hypocycloid,

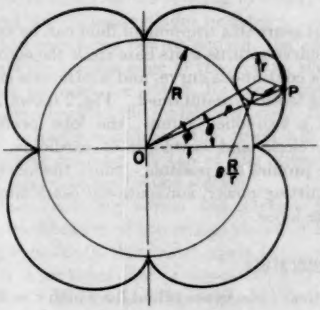


Fig. 1(a) Epicycloid

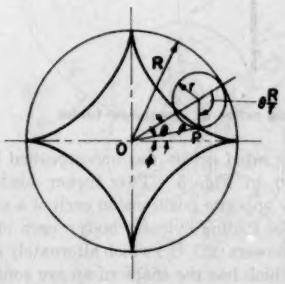


Fig. 1(b) Hypocycloid

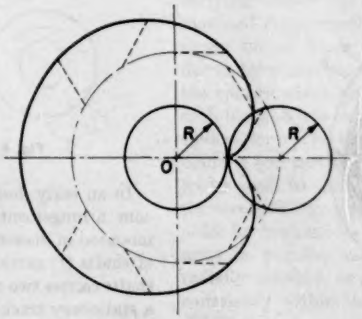


Fig. 1(c) Cardioid

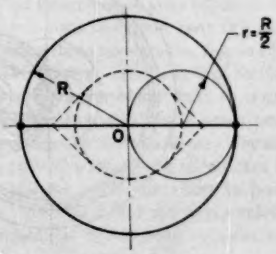


Fig. 1(d) Hypocycloid for $r = R/2$

Contributed by the Machine Design Division and presented at the Annual Meeting, Atlantic City, N. J., November 29–December 4, 1959, of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS.

NOTE: Statements and opinions advanced in papers are to be understood as individual expressions of their authors and not those of the Society. Manuscript received at ASME Headquarters, July 10, 1959. Paper No. 59—A-134.

$$x = 0, \quad y = R \cos \theta$$

which is the equation of a straight line of length $2R$. If the generating circle rolls at uniform angular velocity every point on its periphery performs a rectilinear harmonic motion with the radius of the rolling circle as amplitude.

Another significant cycloidal curve is the epitrochoid which is generated by a point inside the periphery of the rolling circle rolling on the outside of the base circle, which will be discussed later.

Lobar Pumps

Probably the best known application of cycloidal curves in engineering is cycloidal gearing. Gearing of this kind today is, however, not used for the transmission of power but for the design of vanes of lobar pumps which are used as superchargers and certain types of blowers of which the best known is the Roots' blower. A gear pump with involute gear profile has the disadvantage that, because of the discontinuity of the active tooth profiles, fluid is trapped in the space between the tip of one tooth of one gear and the root of the engaging teeth of the other gear.

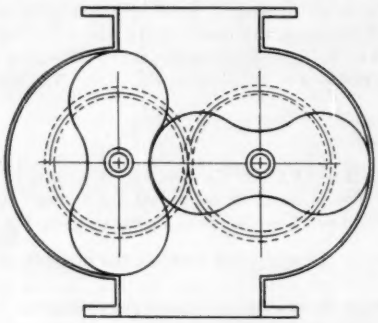


Fig. 2 Lobar pump

With cycloidal gears this trapping of fluid can be avoided. Since the cycloidal curve returns to its base circle the contour of all gear teeth can be a continuous curve, and continuous engagement between meshing teeth be maintained. Fig. 2 shows, as an example, a diagram of a two-lobed pump; the lobe profiles consist of portions of a circle and envelopes of cardioids. Many variations of these profiles are possible. Since the lobes are not suitable for transmitting power, conventional gears must be provided for driving the lobes.

Function Generators

The properties of the hypocycloid for which $r = R/2$ have been used in the design of a harmonic function generator and in computer elements, Fig. 3. A planetary gear (a) is freely rotatable on a crank (b) and engages a stationary ring gear (c) whose pitch diameter is twice as large as its own. The output motion is picked up from a pin (d) attached to the planet gear whose center coincides with a point on the pitch diameter of the planet gear.

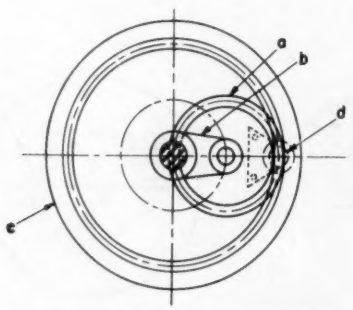


Fig. 3 Harmonic-function generator

Newspaper Folders

Mechanisms based on cycloidal motions have been applied to the folding operation in paper-converting machines.

In a newspaper folder the webs from which the newspaper is made up, after being printing, are assembled and folded lengthwise; the last operation which this folder performs is severing the individual copies from the webs by a crosscut and folding them across. This operation is performed by a folding cylinder (a) Fig. 4, in conjunction with a cutting cylinder (b) and folding rolls (c). The folding cylinder is equipped with several sets of pins (d) which consecutively impale the web near its leading edge and hold it until the leading copy is severed from the remainder of the web. Immediately subsequent to this cutting operation a tucker blade (e) is thrust out of the body of the folding cylinder which forces the center portion of the cutoff into the nip of the rapidly rotating folding rollers while, simultaneously, the pins at the leading edge of the cutoff are withdrawn through cam action. The cutoff released by the folding cylinder is folded across by the action of the folding rollers and dropped on a conveyer. The action of the tucker blade must be performed extremely rapidly since it must take place during the short period when the center of the cutoff passes the nip point of the folding rollers.

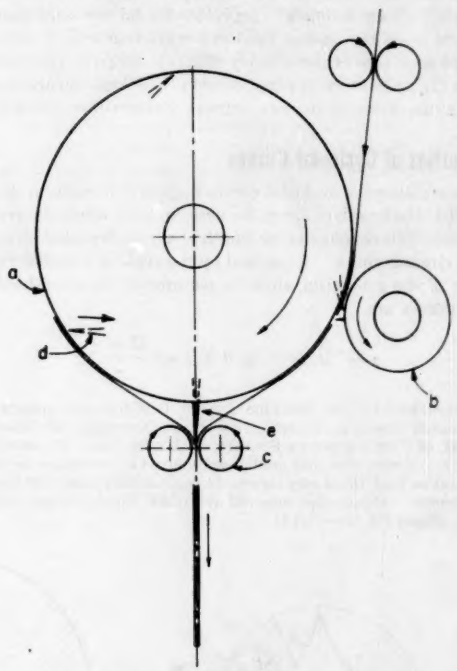


Fig. 4 Folding action of newspaper folder

In an early design this rapid action was accomplished by the cam arrangement shown in Fig. 5. Two tucker blades are mounted in diametrically opposite positions on each of a number of shafts (f) carried in the folding cylinder body; each of these shafts carries two cam followers (m), (n) which alternately engage a stationary track (h), which has the shape of an arc concentric to the cylinder, and a slot (j) on a disk (k), which is arranged in the

area of the folding station and rotates at the same angular velocity as the folding cylinder, but in opposite direction. During that portion of the cycle when both followers of a shaft (f) engage the track (h) the tucker blades are held inside the periphery of the cylinder. When the trailing cam follower (m) reaches the end of the track (h), at position f′, of shaft (f), the leading cam follower (n) enters the slot (j) of disk (k). This causes rapid rotation of the shaft (f) whereby the leading tucker blade is thrust outward at the moment the center of the cutoff passes the folding rollers. After rotation of the shaft (f) through 180 deg relative to the folding cylinder, the cam follower (m) re-enters the track (h) while, simultaneously, follower (n) leaves the slot (j). The other tucker blade, on the same shaft, will go into action during the following revolution of the folding cylinder.

When press speeds were increased the large inertia forces coming into play in this cam action made this arrangement unsatisfactory and a different design was substituted in which the tucker blades perform continuous hypocyclic motions. The hypocycloidal path of the edge of the tucker blade is so selected that only the tip of its cusp protrudes over the periphery of the folding cylinder. This results in an extremely rapid folding action; the hypocycloidal path of the tucker blade is particularly well suited for this operation because the blade edge moves radially at the tip of the cusp which results in a sharp fold.

Fig. 6 shows a diagram of this design. Three tucker blade shafts (f) are provided which carry blades (e) whose length (r) is one half the distance C_0 of the center of the shaft (f) from the center of the folding cylinder. Each shaft (f) is driven by a gear (p) which engages, through an intermediate gear (q), a stationary sun gear (s) whose pitch diameter is three times the pitch diameter of the gear (p), so that gear (p) performs three revolutions, relative to the cylinder, during each revolution of the latter, and a three-cusped hypocycloid is generated. Consecutive positions of the center of the shaft and the tucker blade (shown in dotted lines) between two cusps are marked (f_1), (e_1), (f_2), (e_2), and so on.

Assuming that the folding action begins at the moment that the tucker blade shaft is 25 deg away from the cusp, the polar coordinates of the blade tip at this position, from equations (1) and (2), are $\rho = (2.45)\ r$, $\phi = 1$ deg 40 min; thus, while the tip rotates through this small angle when approaching the cusp it moves radially through more than one half of the radial length of the blade.

In newspaper work it is often required that the folding unit,

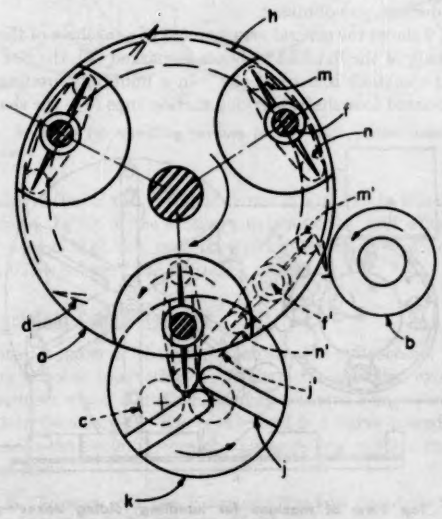


Fig. 5 Cam control for tucker-blade action

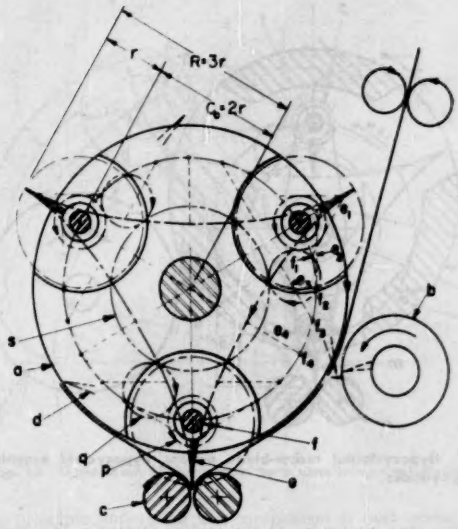


Fig. 6 Hypocycloidal tucker-blade action; hypocycloid concentric with folding cylinder

in addition to folding and cutting, also performs the function of collecting. The collecting operation is necessary if the copy is printed on two consecutive cutoffs; through the collecting operation these cutoffs are placed on top of one another before cross-folding takes place. In a folder arrangement like that shown in Fig. 6 this can be accomplished in the following manner: The cutting cylinder is so proportioned that its circumference is equal to the length of one cutoff (the folding cylinder circumference is equal to three times the length of the cutoff), and it is also provided with a set of pins. The leading edge of every other cutoff is picked up by these pins, and during the ensuing continued rotation of the cutting cylinder, the cutoff is wrapped around the cutting-cylinder periphery. After the cutting cylinder has completed one revolution, the leading edge of the cutoff meets the leading edge of the web on the folding cylinder. At this moment the cutoff, by cam-controlled pin action, is transferred back to the folding cylinder; as the rotation of the folding cylinder continues, the cutoff will be wrapped around the folding cylinder, on top of the subsequent cutoff, and be carried with it to the folding station.

The described collecting action has several disadvantages which led to a modification of this design shown in Fig. 7.[1]

The folding cylinder of this folder again has a circumference whose length is three times that of the cutoff, and it also carries three sets of pins. The body of the folding cylinder is hollowed out and an eccentric shaft (v) passes through its journals (t) carrying spiders (u) on which two tucker-blade shafts (f) are mounted diametrically opposite to one another, each carrying a single tucker blade (m). The tucker-blade shafts are again driven by planet gears (p) and intermediate gear (q) mounted in the spiders which mesh with a stationary sun gear (s), so that each tucker-blade shaft performs three revolutions, relative to the spiders, for one revolution of the latter. The length of the tucker blade is one half the distance of tucker-blade shaft to center of spider shaft so that the edge of the tucker blade again traces a three-cusped hypocycloid relative to the stationary frame; in order for both blades to describe the same path one tucker blade must be pointing radially inward when the other is pointing radially outward, as shown. In this folder the hypocycloid lies completely within the periphery of the folding cylinder except

[1] W. Scott, U. S. Patent 952,714.

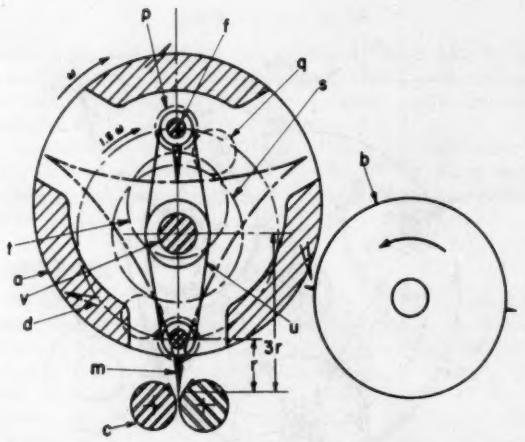


Fig. 7 Hypocycloidal tucker-blade action; hypocycloid eccentric to folding cylinder

that one cusp protrudes at the folding station. The spider is driven at one and one half times the angular speed of the folding cylinder which is the reason why only two folding blades are necessary; when the cylinder rotates through 120 deg—which corresponds to the length of one cutoff—the spider rotates through 180 deg so that, in a noncollect run, each tucker blade will fold off every other cutoff.

This type of folder has the advantage that it permits collecting on the folding cylinder without transfer of a cut-off to the cutting cylinder. For a collect run, one tucker blade is silenced; this can be accomplished by making its drive ineffective and locking it to the spider in the position where it is pointing radially inward. Also, the pin action must be adjusted so that only every other set of pins is withdrawn at the folding station. The consequence will be that every other cutoff will be carried around by the folding cylinder for one revolution without being folded off and will return to the point where the pins on which it hangs first impaled the web; thereby it will get under the web and be positioned under a mating cutoff by virtue of the fact that the length of the circumference of the folding cylinder is an uneven multiple of the length of a cutoff. Since the active tucker blade protrudes from the cylinder only at the folding station it will not disturb the cutoffs during their travel around the cylinder.

In Fig. 8 four consecutive phases a, b, c, and d of this collecting action of the folder are illustrated; the phases are one third of one revolution of the folding cylinder apart (the corresponding rotations of the spider are 180 deg). P_1, P_2, and P_3 are the three sets of pins of the cylinder; mating cutoffs are marked A, B. It is seen that after every 240 deg of rotation of the folding cylinder the mechanism shows the same configuration, except that at phase b the pins P_1 and at phase d the pins P_2 are withdrawn. After another 240 deg of rotation of the folding cylinder the pins P_3 are withdrawn; thus the pin-motion cycle is completed after two revolutions of the folding cylinder.

Machines for Handling of Folding Boxes

Another interesting application of a cycloidal motion is found in a mechanism for preparing folding boxes for packaging. In this mechanism use is made of a cycloidal motion for producing first a nearly rectilinear and then a curvilinear motion of the folding box.[2]

Folding boxes are shipped in flat shape. Before use they must

[2] M. Hornsteiner, "Bebriebetschnische Fragen bei der Faltschachtel-Fertigung," VDI-Z99, 1957, pp. 52–53.

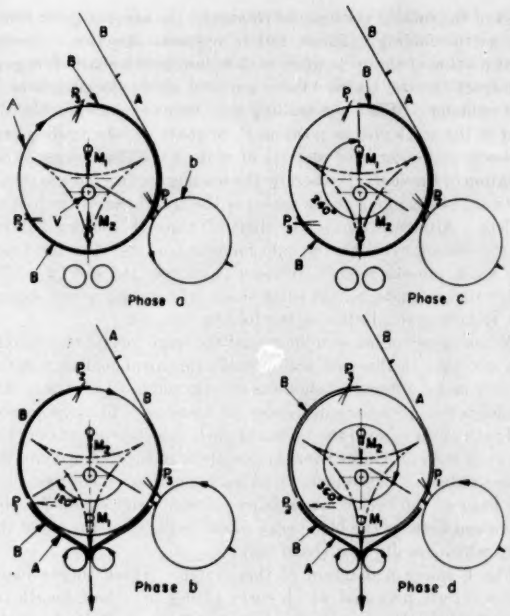


Fig. 8 Collecting action of folder with hypocycloid eccentric to folding cylinder

be unfolded, and then folded in the opposite direction in order to remove the "set" which they acquire during storage. This is accomplished by picking up a flat folding box from the face of a stack by means of suction cups, moving the box in a curvilinear path so that one edge of the box slides along a cam, and passing it through a pair of rollers. In older machines the motion of these suction cups was performed by a reciprocating arm; because the return stroke of this arm was idle the rate of production of this machine was low. In a newer design of machines for this purpose a multiplicity of suction cups is provided which perform a continuous cycloidal motion. While one suction cup picks up the flat folding box, unfolds it, and transfers it to rollers and a conveyer, the following suction cup already approaches the stack and initiates the new cycle; by this action a great increase in the rate of production was obtained.

Fig. 9 shows the general arrangement of a machine of this kind. The stack of the flat folding boxes is marked (a), the box at the face of the stack is marked (b). In a uniformly rotating wheel are mounted four shafts carrying suction cups (c); the shafts are

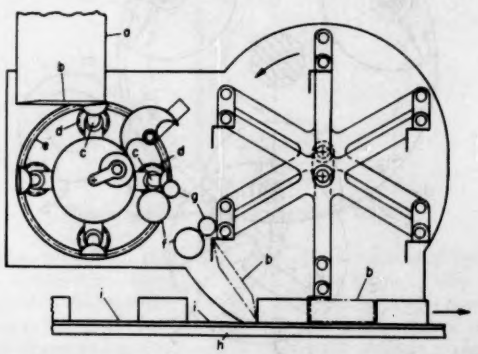


Fig. 9 Top view of machine for handling folding boxes—general arrangement

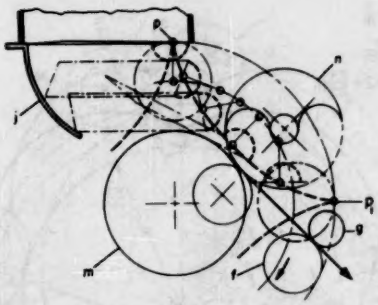


Fig. 10 Machine for handling folding boxes; hypocycloidal motion of suction cups

connected to planet gears (d) engaging a stationary ring gear (e) so that the tips of the suction cups describe a four-cusped hypocycloid. At one cusp of this path the tip of the suction cup comes in contact with the box (b); as the suction cup continues to move it pulls the box (b) away from the stack. This brings one edge of the box in contact with the cam surface (j), Fig. 10, and sliding of the edge along this cam causes the box to unfold. As the suction cups approach the apex of the hypocycloid between the cusps (p), (p_1) the box comes in contact with roller (m) whereby the box is folded in the opposite direction. At the apex between (p) and (p_1), a roller (n) with an interrupted circumference engages roller (m); these rollers complete the folding of the boxes in the opposite direction and pass them on to two other pairs of flattening rollers (g), (f). From these rollers the boxes move on to a conveyer (h) with pockets (i); when their front edges strike the crossbar forming the front end of the pockets their rear edges come in contact with angles mounted on a Ferris-wheel type rotating structure (q), which pushes the boxes completely into the pockets of the conveyer. The tangential velocity of these angles is equal to the speed of the conveyer.

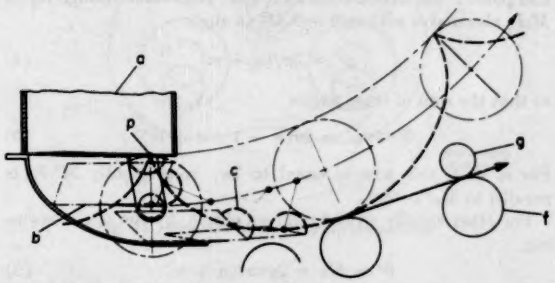


Fig. 11 Machine for handling folding boxes; epicycloidal motion of suction cups

An alternative of this design is shown in Fig. 11; in this design the path of the tip of the suction cup is an epicycloid which has the advantage that this path does not intersect the path of the folding boxes between the flattening roller.

The Cycloidal Speed Reducer

Another member of the cycloid family, the epitrochoid, is the basis for another interesting mechanism, the so-called cylcoidal speed reducer which is distinguished by extreme compactness (it is built in Germany).[3] The epitrochoid is a curve traced by a point inside the circumference of a circle rolling on the outside of the base circle.

[3] D. W. Botsiber and Leo Kingston, "Cycloid Speed Reducer," *Machine Design*, June 28, 1956, pp. 65–69.

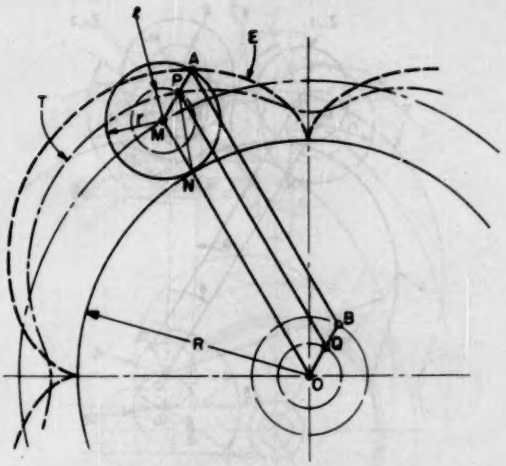


Fig. 12 Generation of epitrochoid by a parallelogram linkage

The principle underlying this mechanism is best understood by considering the means by which an epicycloid E may be generated. In a rolling circle, Fig. 12, the radius of length r of a point A on its periphery may be represented by a link connecting A with its center M. Since during the rolling motion the distance between M and the center O of the base circle—with radius R—remains unchanged, this distance may be represented by a link M-O. If a link O-B of length r and a link A-B of length O-M is added it is obvious that links O-B and A-B remain parallel to M-P and O-M, respectively, as the circle M, r rolls on the base circle. Links O-B, M-A, and links O-M and A-B, respectively, must therefore have equal angular velocity during the rolling motion; furthermore, the ratio between their angular velocities must be constant. Consequently, the epicycloid E can be generated by the parallelogram linkage O-M-A-B if the links O-M, O-B are driven at angular velocities which have a fixed ratio determined by their geometric proportions.[3] If a point P on the link M-A is connected with a point Q on the link O-B, chosen so that distance O-Q is equal to distance M-$P(=\rho_0)$, through a link of length O-M, link P-Q will rotate with the same angular velocity as the link O-M. If point P is inside the periphery of the rolling circle it traces a curve T called an epitrochoid which has $n = R/r$ lobes. Since the epitrochoid has no cusps it is suitable for continuous cam action if n is an integer number.

If a roller Z of radius z be attached to point P and participated in its motion along the epitrochoidal path the envelope of all roller positions will be the equidistant curve C of the epitrochoid, Fig. 13. The point of contact of the roller and the curve must lie on line P-N where point N on M-O, the point of tangency of the generating circle and the base circle, is the instantaneous center of the motion of the generating circle.

If the center of the generating circle rotates counterclockwise, about O, through an angle θ, from M_0 to M_1, and point P, with roller Z, moves from position P_0, Z_0 to position P_1, Z_1, links M-P and O-Q will rotate, relative to O-M, through an angle $\theta R/r$.

We now perform a kinematic inversion: Roller Z is held stationary at Z_0, and the disk with contour C is moved so that it performs the inverse of the motion by which the contour was generated, thus maintaining contact with the roller. The generating motion of the roller Z was the compound motion of P on link Q-P from position Q_0-P_0 to Q_1-P_1; this motion consisted of the rotation of pivot Q, on Q_0-O, about O through angle $\theta R/r$, and rotation of link Q-P through angle θ, both counterclockwise. The inverse of this motion, if link Q-P with roller Z is

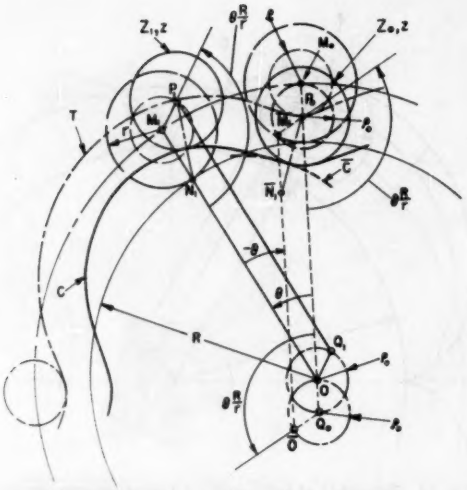


Fig. 13 Kinematic relation between epitrochoidal cam and parallelogram linkage

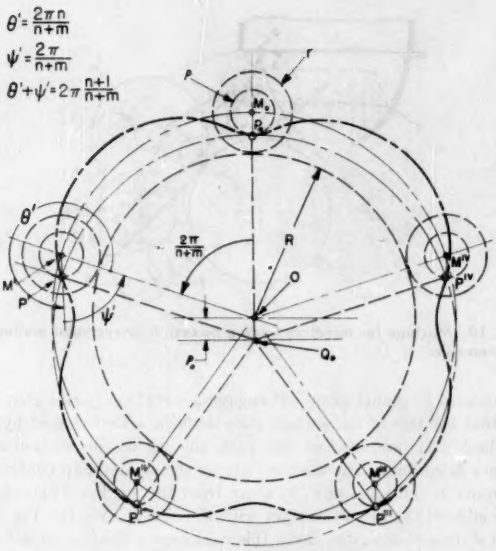


Fig. 14 Relation between epitrochoid and regular polygon

held stationary at Q_0-P_0, is obtained by rotating pivot O (the center of the disk) on link O-Q_0 about Q_0, through the angle $\theta R/r$, and then rotating the disk about its center O, through angle θ, both clockwise. The corresponding link configuration at the end of the inverse motion is indicated by the dotted lines Q_0-$\bar{O}$-$\bar{M}_1$-P_0; it is seen that it is congruent to Q_1-O-M_1-P_1. The disk contour, at the same instant, is in position C; the instantaneous center is $\bar{N}_1$ on link $\bar{O}_1$-M_1 its position corresponding to N_1. In the compound motion of the disk the ratio of the angular speed of the disk center O about Q_0 to that of the disk about its center O is R/r.

A mechanism with these elements—a stationary roller, an output member consisting of a disk with curve C as periphery, and an input member consisting of a crank of length Q-O—appears to be capable of large reduction of speed since the speed ratio depends only on the ratio of R/r, and the radius r can be made quite small. However, if the mechanism would contain only these elements it would not yet be operable. One reason is that the output motion of the disk is not a pure rotation but a compound motion. The motion superposed to its rotation is a translation in a circular path whose radius is equal to the length of crank Q-O, and the speed of its center is equal to the speed of this crank. This translation can be filtered out by connecting the disk C to the output member not directly but by cranks of length ρ_0; instead, pins may be used which engage circular holes in the output member whose diameter is equal to the pin diameter plus the diameter $2\rho_0$ of the crank circle. The motion of these pins, relative to the center of the holes which they engage, must be in phase with that of the crank Q-O. Through this arrangement the translation in a circular path of the disk relative to the output member is possible, while only the rotary motion of the disk is transmitted to this member. The pins and holes can be interchanged such that the holes are provided in the disk, and the pins in the output member without changing the kinematical relation (Fig. 16).

Another reason why the discussed mechanism is not yet operable is that the crank Q-O and one roller Z do not sufficiently constrain the disk to remain in contact with the roller. For complete constraint additional stationary rollers are needed. The location of these rollers can be determined if another interesting property of epicyclic curves is considered.

Attention is focused, in an epicyclic curve with n lobes, on the relation of points of this curve which are associated with equally spaced radii of the base circle. The angle of this spacing is denoted by $2\pi/(n + m)$ where m is an integer number, Fig. 14. When, starting at the cusp, the radius O-M has moved to position O-M' through angle $2\pi/(n + m)$ the rolling circle with radius r has rotated, relative to O-M', through an angle

$$\theta' = 2\pi n/(n + m) \tag{3}$$

and point P has moved from P_0 to P'. A parallel through M' to M_0P_0 also makes with radius O-M' an angle

$$\psi' = 2\pi/(n + m) \tag{4}$$

so that the sum of these angles

$$\theta' + \psi' = 2\pi(n + 1)/(n + m) \tag{5}$$

For $m = 1$, this sum is equal to 2π; consequently M'-P' is parallel to M_0P_0.

For other equally spaced radii we obtain, by the same reasoning,

$$\theta^x = \theta' x = 2\pi nx/(n + m) \tag{6}$$

$$\psi^x = 2\pi x/(n + m) \tag{7}$$

$$\theta^x + \psi^x = 2\pi x(n + 1)/(n + m) \tag{8}$$

Again, for $m = 1$, and for any integer value of x, this angle will be a multiple of 2π; therefore the radii M^xP^x will be parallel to M_0P_0 in the specified positions. For this reason, since the centers M_0, M', M^x are lying on the circle O, $(R + r)$, the points P, $P' \ldots P^x$ of the epitrochoid must lie on a circle with the same radius whose center is a point Q_0, at a distance ρ_0 from O, in the direction of M_0-P_0. The number of points of the epitrochoid lying on this circle is by one larger than the number of lobes of the epitrochoid.

We next consider point P_1, Fig. 15, which is associated with the radius O-M_1 which makes an angle θ with the radius O-M_0 at the cusp, and point P_2 associated with radius O-M_2 making an angle $2\pi/(n + 1)$ with O-M_1. Applying the same reasoning as previously it is seen that M_2-$P_2 \ldots M_x$-P_x will again be

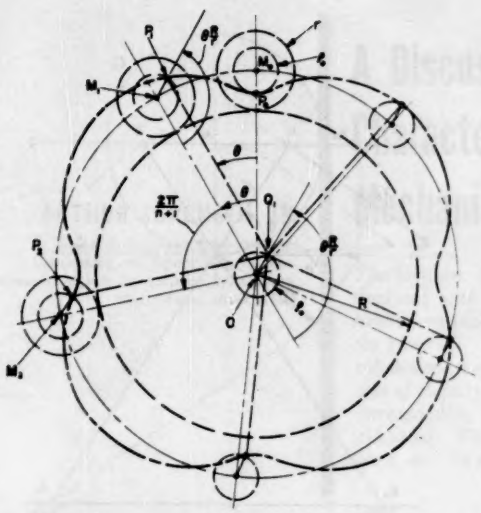


Fig. 15 Relation between epitrochoid with n lobes and moving regular polygon with $(n+1)$ corners

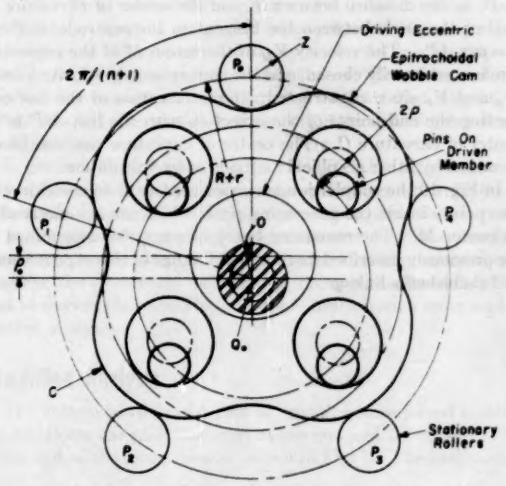


Fig. 16 Diagram of cycloidal speed reducer

parallel to M_1-P_1, that the points P_1, P_2 . . . . must lie again on a circle whose center Q_1 is at distance ρ_0 from the center O of the base circle parallel to M_1-P_1 . . . . M_x-P_x. Since this relation applies to any angle θ it must be concluded that any group of $(n+1)$ points on an epitrochoid with n lobes associated with $(n+1)$ equally spaced radii of the base circle form equally shaped regular polygons inscribed in a circle with radius $(R+r)$; that the center Q_1, Q_2 . . . . of these polygons lies on a circle with radius O-Q which is equal in length and parallel to the radius ρ_0 of the generating circle. If one such polygon moves such that all its corners trace the epitrochoid, the radii P_1-Q_1, P_2-Q_1 . . . . remain in a fixed relationship to one another.

If $(n+1)$ rollers Z_1 . . . . of radius z are placed at points P_1 . . . . of the n-lobed epicycloid associated with radii of the base circle which are $2\pi/(n+1)$ radians apart the points P_1 . . . . will form a regular polygon with center Q_1, and these rollers will be in contact with the equidistant curve C of the epitrochoid. This polygon contains the link P_1-Q_1; therefore, when the epitrochoid is generated, the polygon performs the same motion as this link. In the kinematic inversion previously discussed this link remained fixed. From this it is concluded that, if the whole polygon remains fixed, contact between the five rollers on the polygon and the curve C will be maintained if the curve disk performs the compound motion previously described. Since these rollers engage opposite flanks of the lobes of the periphery of the disk C the disk will be adequately constrained.

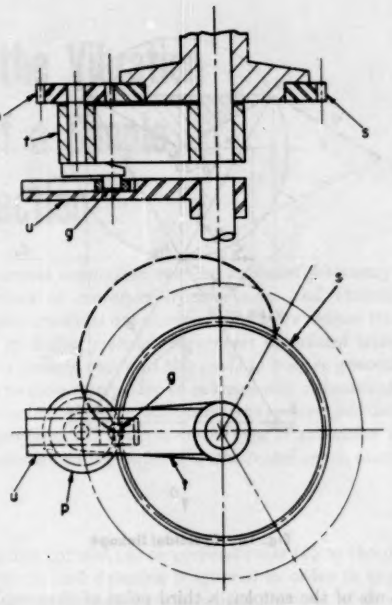


Fig. 17 Cycloidal crank

By placing equally spaced $(n+1)$ rollers on a circle with center Q_0 and radius $(R+r)$ the cycloidal speed reducer is completed, as indicated in Fig. 16. An eccentric with radius ρ_0 on the input shaft constitutes the crank Q_0-O. Input and output shaft are concentric; the ratio is R/r.

Cycloidal Crank

In an epicyclic gear train consisting of a fixed sun gear (s), a driving spider (t), and a planet gear (p), Fig. 17, a crank (g) is attached to the planet gear at its pitch line; this crank engages a shoe in a slotted crank (u) of the driven shaft which shaft will then perform a cyclic, nonuniform motion whose acceleration and velocity at the cusps of the cycloidal path described by the crankpin center will be equal to zero. It is possible to obtain dwells at the cusps by the use of interrupted gears.[4]

Cycloidal Linkages

Eventually the so-called cycloidal linkages may be mentioned. These are four-bar linkages which are used for generating an approximately straight-line motion of a point of the coupler. The cycloidal path which is traced by each point of a circle rolling in a plane along a straight line degenerates, for the center of the circle, into a straight line. Vice versa, if two points of the generating circle move in correlated cycloidal paths the center of the circle will describe a straight line. Therefore, if the lengths of the two cranks to which the coupler of a four-bar linkage is connected are equal to the radii of curvature of the cycloids described by those points of the generating circle which correspond

[4] E. H. Schmidt, "Cycloidal Crank Mechanisms," *Machine Design*, April 2, 1959, pp. 111–114.

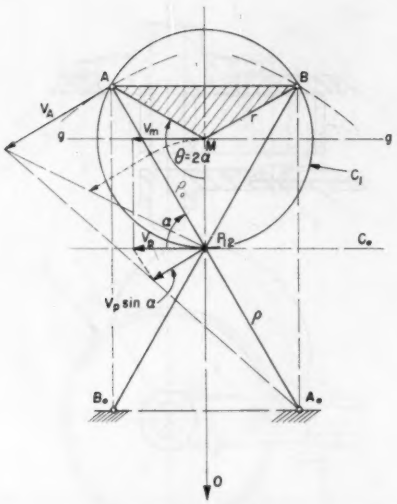


Fig. 18 Cycloidal linkage

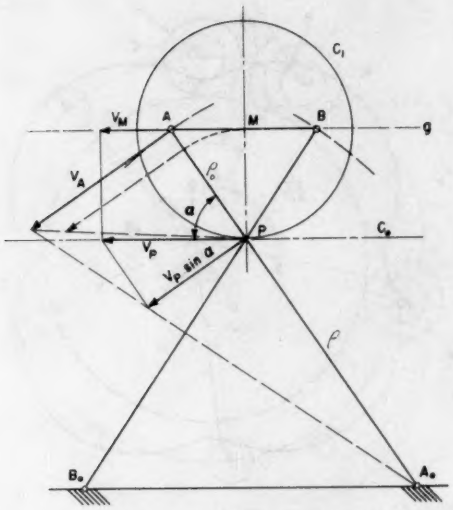


Fig. 19 Tchebisheff's linkage

to the joints of the coupler, a third point of the coupler, which corresponds to the center of the generating circle, will perform an approximately straight-line motion.

In Fig. 18 two points on the periphery of the generating circle are selected as the joints of the coupler. The centers of curvature may be determined by Euler-Savary's theorem[5]:

$$V_A = V_P(\sin \alpha)(\rho_0 + \rho)/\rho$$

where V_A is the velocity of point A of the moving plane, V_P the velocity of change of the instantaneous center P, ρ the distance A-P, ρ_0 the distance between P and the center of curvature C_0, and α, the angle between the tangent to the centrode at P and the ray AP. The velocity V_M of the center M of the generating circle is arbitrarily chosen, and the corresponding velocity vectors V_A and $V_P \sin \alpha$ constructed; the intersection of the line connecting the end points of these vectors with the line A-P is the center of curvature C. The center of curvature can also be determined by other graphical methods or by calculation.

In Fig. 19 the coupler joints are selected so as to coincide with two points inside the generating circle which are symmetrical to its center M. The remaining two joints may be determined by the previously described method. A linkage of this type is known as Tchebisheff's linkage.

[5] L. Burmester, "Kinematik," Arthur Felix, Leipzig, Germany, 1888, p. 125.

A Discussion of the Vibration Characteristics of a Simple Mechanical Connection

ARTHUR SORENSEN, JR.
Research Engineer, Environmental Laboratory, AC Spark Plug Division, General Motors Corporation, Milwaukee, Wis. Assoc. Mem. ASME

The behavior of a simple mechanical connection over an extended frequency range is analyzed with particular reference to vibration transmission and isolation. The limitations inherent in a simplified analysis are clarified by a more refined treatment of the problem. The significance of higher critical frequencies associated with the distributed mass of the connection is pointed out, and this leads to a more generalized concept of rigidity. A qualified definition of rigidity is subsequently introduced, and the corresponding performance characteristics of a vibration transmitter and isolator are discussed. The dependency of performance criteria on the type of excitation is emphasized, and the essential ideas embodied in the paper are illustrated in an example.

Introduction

IT IS FREQUENTLY NECESSARY to interconnect components in a vibrating system by means of a suitable connection. The nature of the connection will depend on the relationship desired between the vibration amplitudes of the connected elements. For example, a very high degree of rigidity may be required when it is essential to maintain a prescribed geometrical configuration.

In general, the behavior of physical systems is extremely complicated, and certain simplifying assumptions are unavoidable in a theoretical analysis. Although these assumptions may be valid in many situations, in some circumstances they may lead to erroneous conclusions. It is the purpose of this presentation to examine the elementary behavior of a simple mechanical system and to discuss the modifications which result from a more sophisticated analysis.

Simplified Analysis

The system consists of a pair of "rigid" bodies joined together by a uniform bar with the input at one end and the output at the other end of the connection as shown in Fig. 1. The excitation is applied either parallel (a) or perpendicular (b) to the center line of the system, and damping is ignored in order to expedite the analysis and bring out certain salient features of the idealized system. In the lateral vibration, the rotation of the mass at the input end is assumed to be negligible, and the connection may be regarded as a cantilever with one end built into a sliding support. The output response is specified for both the longitudinal and lateral modes of vibration, and a determination of the corresponding input motion is required.

In the vibration analysis [1][1] of mechanical systems, it is customary to think in terms of lumped masses and massless springs. On this basis, the relationship between the input and output may be established in a straight-forward manner by writing Newton's second law of motion for the mass at the output end:

$$K(\lambda_1 - \lambda_2) = \frac{W_2}{g}\ddot{\lambda}_2 \qquad K = \frac{F_0}{\Delta} \tag{1}$$

where λ represents either u or v provided the proper K is used. The acceleration is obtained by differentiation as follows:

$$\lambda_2 = a \sin \omega t \qquad \ddot{\lambda}_2 = -\omega^2 \lambda_2 \tag{2}$$

where a dot indicates differentiation with respect to time. A substitution of (2) into (1) then yields:

$$\lambda_1 = \left(1 - \frac{W_2\omega^2}{gK}\right)\lambda_2 \tag{3}$$

Contributed by the Machine Design Division and presented at the Annual Meeting, Atlantic City, N. J., November 29–December 4, 1959, of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS.

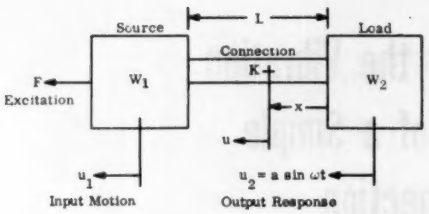
NOTE: Statements and opinions advanced in papers are to be understood as individual expressions of their authors and not those of the Society. Manuscript received at ASME Headquarters, June 11, 1959. Paper No. 59—A-41.

[1] Numbers in brackets designate References at end of paper.

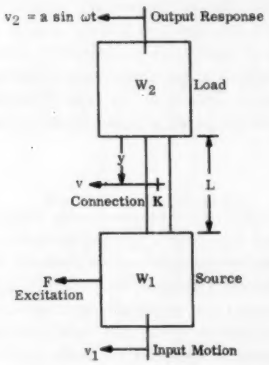
Nomenclature

A = cross-sectional area of connection, sq in.
a = vibration amplitude of load, in.
B = excitation bandwidth, cps
b = width of connection, in.
C_n = coefficients, where n = 1, 2, 3, ...
c, k = longitudinal and lateral system parameters
d = thickness of connection, in.
E = modulus of elasticity, psi
F = instantaneous excitation force, lb
F_0 = static load, lb
f = excitation frequency, cps
f_a, f_b = lower and upper boundary of excitation frequency range, cps
f_2 = critical frequency, cps
G = mean square acceleration density, g^2 per cps
g = acceleration due to gravity, in. per sec sq
I = transverse area moment of inertia, in.^{4}
J = transmissibility integral, cps
K = stiffness of connection, lb per in.
L = length of connection, in.
p, q = longitudinal and lateral frequency parameters
Q = rms transmissibility
R = relative amplitude
r = frequency ratio
s = weight ratio

(*Continued on next page*)



(a) Longitudinal Vibration



(b) Lateral Vibration

Fig. 1 Idealized mechanical system

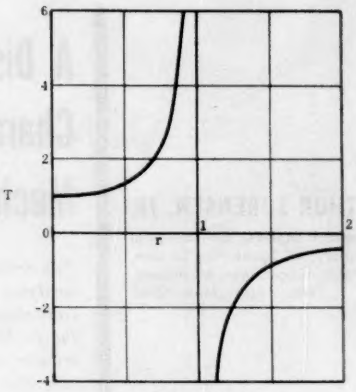
where the stiffness [2] for the longitudinal and lateral modes of vibration is given by:

$$K = \frac{EA}{L} = \frac{Ebd}{L} \text{ (longitudinal)} \tag{4a}$$

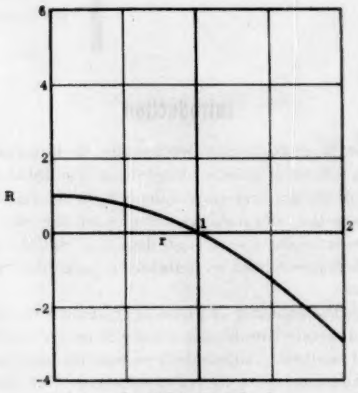
$$K = \frac{3EI}{L^3} = \frac{Ebd^3}{4L^3} \text{ (lateral)} \tag{4b}$$

in which the more specific expressions pertain to a rectangular section. It is immediately apparent that for a relatively slender bar, a beam type of connection is inherently more flexible than a prism since the stiffness of the former varies with the cube of the d/L ratio while the latter is directly proportional to this quantity.

The amplitude relation may be put into nondimensional form as follows:



(a) Transmissibility



(b) Relative Amplitude

Fig. 2 Amplitude variation as a function of frequency ratio

$$T = \frac{\lambda_2}{\lambda_1} = \frac{1}{R} \qquad R = 1 - r^2 \tag{5}$$

where

$$r = \frac{\omega}{\omega_2} = \frac{f}{f_2} \qquad \omega = 2\pi f \qquad \omega_2 = 2\pi f_2 = \sqrt{\frac{\varrho K}{W_2}} \tag{6}$$

The alternative representations in terms of the transmissibility T or relative amplitude R are displayed as a function of the frequency ratio r in Fig. 2. Although the former method of presentation is usually employed, the latter has certain advantages which will become apparent in the ensuing analysis.

Nomenclature

T = transmissibility
$T(t)$ = a function of time
t = time, sec
u, v = longitudinal and lateral displacement of point on uniform bar, in.
u_1, v_1 = longitudinal and lateral displacement of input end of connection, in.
u_2, v_2 = longitudinal and lateral displacement of output end of connection, in.
W_1, W_2 = weight of source and load, lb
w = weight of connection, lb
$X(x), Y(y)$ = a function of position in the longitudinal and lateral modes
x, y = distance of point on uniform bar from output end of connection in longitudinal and lateral modes, in.
α, β = constants
γ = unit weight of connection, lb per cu in.
Δ = static deflection, in.
θ = angle
λ = generalized displacement of point on uniform bar, in.
λ_1, λ_2 = generalized displacement of input and output end of connection, in.
σ = unit stress due to steady load, psi
ω = circular frequency, rad per sec

An examination of the curves indicates that the relationship between the input and output amplitude depends entirely on the ratio of the excitation frequency to the critical frequency f_2. If the frequency ratio is small compared to unity, the connection may be considered rigid, and the entire system then vibrates in a relatively uniform manner. At the other extreme, when the frequency ratio is substantially greater than unity, the input motion is not communicated to the output end of the connection, and the latter becomes, in effect, a vibration isolator [3].

The subsequent discussion will be concerned with two particular situations in which the connection is intended to serve either as (1) a rigid means of vibration transmission to the load which is required to vibrate at a specified amplitude, or (2) a flexible means of vibration isolation from the source which vibrates at a designated amplitude.

Apparently, a high frequency ratio will oppose the former and favor the latter objective. Hence it is sufficient to adjust the frequency ratio either above or below unity, respectively, when isolation or transmission is desired. Unfortunately, in practice the matter is not quite so simple, and in spite of a desire to provide a rigid connection, operation at a high frequency ratio may be unavoidable, resulting in an excessive input amplitude. A more accurate appraisal will show that the outlook is not quite so bleak as it would appear, since a further increase in the frequency ratio will eventually result in an increase in transmissibility. Consequently, at sufficiently high excitation frequencies, the output vibration level will be amplified rather than attenuated. While this is desirable from the viewpoint of vibration transmission, it places a limit on the range of effective isolation.

Refined Analysis

The previous analysis is quite attractive in view of its relative simplicity, and the solution is in many instances a good approximation. Evidently, the result of the simplified analysis is a special, limiting case of a more general solution. This raises some interesting questions concerning (1) the conditions under which the approximation is valid and (2) the behavior of the system when these conditions are not satisfied.

The simplified analysis ignored the mass of the connection. Of course, any physical system is actually composed of a continuous distribution of material. It is therefore pertinent to inquire how the behavior of the system is influenced by the distributed mass of the uniform bar. While it is true that the elasticity of the body at either end of the connection was likewise ignored, situations do arise in which the geometry of these elements justifies this assumption.

Longitudinal Vibration: The partial differential equation of motion for a uniform prism [4] is given as:

$$\frac{\partial^2 u}{\partial t^2} = c^2 \frac{\partial^2 u}{\partial x^2} \qquad \text{where} \quad c = \sqrt{\frac{gE}{\gamma}} \tag{7}$$

It is a standard procedure in partial differential equations [5] to assume a solution of the form:

$$u(x, t) = X(x)T(t) \tag{8}$$

where $X(x)$ is a function of x alone, and $T(t)$ is a function of t alone. A differentiation and substitution into the differential equation of motion leads to the following solution:

$$T(t) = C_1 \sin(\alpha t) + C_2 \cos(\alpha t) \tag{9a}$$

$$X(x) = C_3 \sin\left(\frac{\alpha}{c} x\right) + C_4 \cos\left(\frac{\alpha}{c} x\right) \tag{9b}$$

The applicable boundary conditions for this problem are:

$$u(o, t) = a \sin \omega t \tag{10a}$$

$$EA \frac{\partial u(o, t)}{\partial x} = \frac{W_2}{g} \frac{\partial^2 u(o, t)}{\partial t^2} \tag{10b}$$

From the first of these conditions:

$$\alpha = \omega \qquad C_2 = 0 \qquad C_1 C_4 = a \tag{11a}$$

The second boundary condition then requires that:

$$C_1 C_3 = -\frac{W_2 c a \omega}{gEA} \tag{11b}$$

from which the instantaneous displacement at any point in the bar is given as a function of time as follows:

$$u(x, t) = a\left[\cos\left(\frac{\omega}{c} x\right) - \left(\frac{W_2 c \omega}{gEA}\right) \sin\left(\frac{\omega}{c} x\right)\right] \sin \omega t \tag{12}$$

If the following substitutions are made:

$$p = \frac{\omega L}{c} = \frac{2\pi L f}{c} \quad \text{and} \quad s = \frac{w}{W_2} \quad \text{where} \quad w = \gamma AL = \gamma bdL \tag{13}$$

the solution then becomes:

$$u(x, t) = a\left[\cos\left(\frac{\omega}{c} x\right) - \frac{p}{s} \sin\left(\frac{\omega}{c} x\right)\right] \sin \omega t \tag{14}$$

The relative amplitude is then given by:

$$R = \frac{1}{T} = \frac{u_1}{u_2} = \frac{u(L, t)}{u(o, t)} = \cos p - \frac{p}{s} \sin p \tag{15}$$

This expression indicates the desirability of graphing the relative amplitude in preference to its reciprocal, the transmissibility, which becomes infinite at various frequencies resulting in numerous discontinuities. The relative amplitude R is graphed as a function of the longitudinal frequency parameter p for several values of the weight ratio s in Fig. 3. The reader's attention is directed to the change of scale employed for the sake of expedience in these and subsequent curves.

Lateral Vibration: The partial differential equation for a uniform beam [4] is given as:

$$\frac{\partial^2 v}{\partial t^2} = -k^2 \frac{\partial^4 v}{\partial y^4} \qquad \text{where } k = \sqrt{\frac{gEI}{\gamma A}} = \frac{d}{2}\sqrt{\frac{gE}{3\gamma}} \tag{16}$$

If the standard approach is again employed:

$$v(y, t) = Y(y)T(t) \tag{17}$$

and a repetition of the procedure adopted previously then yields the following solution:

$$T(t) = C_1 \sin(\beta t) + C_2 \cos(\beta t) \tag{18a}$$

$$Y(y) = C_3 \sin\left(\sqrt{\frac{\beta}{k}}\, y\right) + C_4 \cos\left(\sqrt{\frac{\beta}{k}}\, y\right) + C_5 \sinh\left(\sqrt{\frac{\beta}{k}}\, y\right) + C_6 \cosh\left(\sqrt{\frac{\beta}{k}}\, y\right) \tag{18b}$$

The boundary conditions are:

$$v(o, t) = a \sin \omega t \tag{19a}$$

$$\frac{\partial^2 v(o, t)}{\partial y^2} = 0 \tag{19b}$$

$$-EI \frac{\partial^3 v(o, t)}{\partial y^3} = \frac{W_2}{g} \frac{\partial^2 v(o, t)}{\partial t^2} \tag{19c}$$

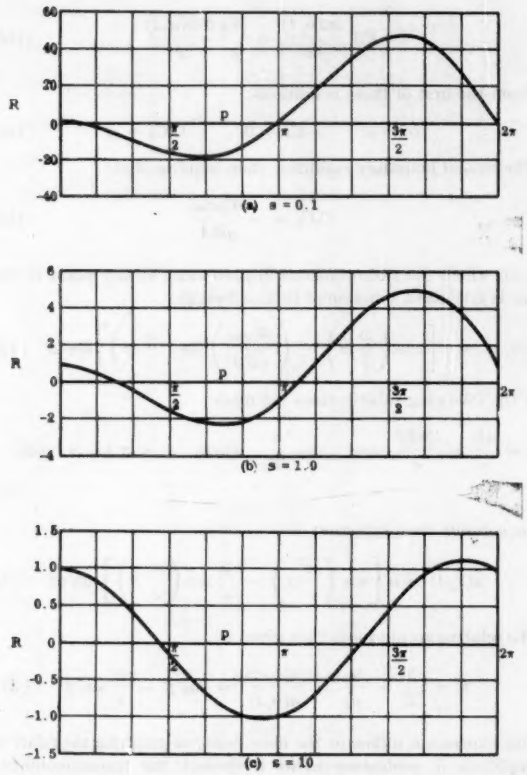


Fig. 3 Relative amplitude as a function of frequency parameter for various weight ratios in longitudinal direction

$$\frac{\partial v(L,t)}{\partial y} = 0 \tag{19d}$$

Successive application of the boundary conditions then yields:

$$\beta = \omega \qquad C_2 = 0 \qquad C_1(C_4 + C_5) = a \tag{20a}$$

$$C_4 = C_5 \tag{20b}$$

$$\frac{C_5 - C_3}{C_4} = \frac{2W_2\sqrt{\omega k^3}}{gEI} \tag{20c}$$

$$\frac{C_3}{C_4} = \frac{\sin\left(L\sqrt{\frac{\omega}{k}}\right) - \sinh\left(L\sqrt{\frac{\omega}{k}}\right) - \left(\frac{2W_2\sqrt{\omega k^3}}{gEI}\right)\cosh\left(L\sqrt{\frac{\omega}{k}}\right)}{\cos\left(L\sqrt{\frac{\omega}{k}}\right) + \cosh\left(L\sqrt{\frac{\omega}{k}}\right)} \tag{20d}$$

The relative amplitude is then given by:

$$R = \frac{1}{T} = \frac{v(L,t)}{v(o,t)}$$

$$= \frac{1 + \cos q \cosh q - \frac{q}{s}(\sin q \cosh q - \cos q \sinh q)}{\cos q + \cosh q} \tag{21}$$

where

$$q = L\sqrt{\frac{\omega}{k}} = L\sqrt{\frac{2\pi f}{k}} \tag{22}$$

The quantity R is graphed as a function of the lateral frequency parameter q for several values of s in Fig. 4.

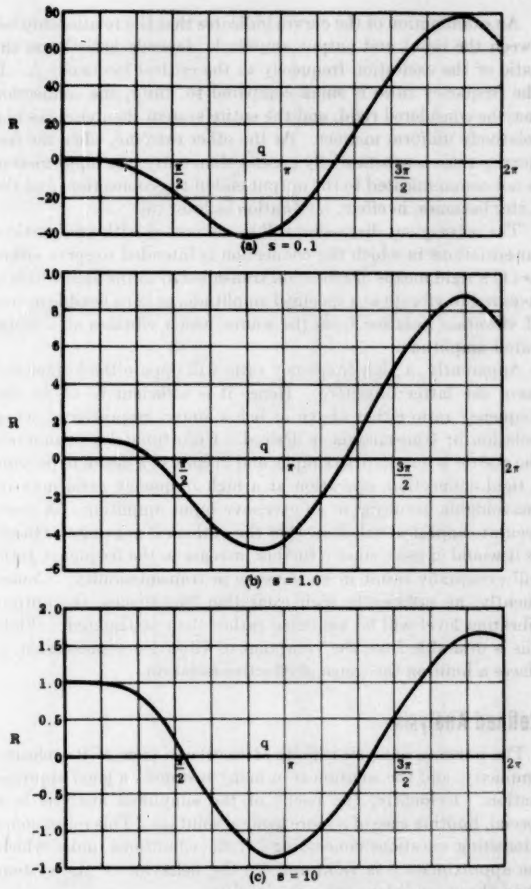


Fig. 4 Relative amplitude as a function of frequency parameter for various weight ratios in lateral direction

Discussion of Results

A comparison of the relative amplitude curve of Fig. 2(*b*) with those of Fig. 3 and 4 reveals a fundamental difference in the basic character of the curves. In both instances, the relative amplitude starts out at unity when the frequency is zero. It decreases steadily to zero with an increase in frequency and thereafter becomes negative. However, in lieu of the former unlimited negative increase with frequency, the relative amplitude is now oscillatory in nature with progressively increasing peaks. This radical alteration is desirable from the standpoint of vibration transmission but means that an upper as well as a lower limit is imposed on the range of effective isolation.

Equivalence of Results: It is easy to demonstrate that the expression obtained from the simplified analysis is a special, limiting case of the more general solution established in the preceding section. Furthermore, equations (15) and (21) reduce to (5) when p and q are relatively small, provided s is not too large:

$$R \approx 1 - \frac{W_2\omega^2}{gK} = 1 - r^2$$

where the appropriate expression for K obtained from (4*a*) or (4*b*) is inserted above. If

$$p < 1 \quad \text{and} \quad q < 1$$

the trigonometric and hyperbolic functions may be approximated by the first two terms of the corresponding Maclaurin expansion:

$$\sin\theta \approx \theta - \frac{\theta^3}{6} \qquad \cos\theta \approx 1 - \frac{\theta^2}{2}$$

$$\sinh\theta \approx \theta + \frac{\theta^3}{6} \qquad \cosh\theta \approx 1 + \frac{\theta^2}{2}$$

The relative amplitude is then given approximately by:

$$R \approx 1 - \frac{p^3}{2} - \frac{p^2}{s}\left(1 - \frac{p^2}{6}\right) \approx 1 - \frac{p^2}{s}$$

$$= 1 - \frac{W_2 L\omega^2}{gEA} = 1 - r^2$$

provided either $p \ll \sqrt{2}$ or $s \ll 2$ (longitudinal)

$$R \approx 1 - \frac{q^4}{8} - \frac{q^4}{3s} \approx 1 - \frac{q^4}{3s} = 1 - \frac{W_2 L^2\omega^2}{3gEI} = 1 - r^2$$

provided either $q \ll \sqrt[4]{8}$ or $s \ll \frac{8}{3}$ (lateral)

Interpretation of Behavior: In the limiting case, when the load attached to the output end of the bar vanishes, the following expressions are obtained:

$$R = \cos p \text{ (longitudinal)} \qquad (23a)$$

$$R = \frac{1 + \cos q \cosh q}{\cos q + \cosh q} \text{ (lateral)} \qquad (23b)$$

These expressions oscillate between plus and minus unity. This represents the optimum condition from the viewpoint of vibration transmission over an unrestricted frequency range in that the peak input amplitude is minimized. Since the transmissibility across the connection is always greater than or equal to unity, it is impossible to achieve vibration isolation at any frequency. Inasmuch as this limiting condition is approached when the weight of the connection increases, we may conclude that a relatively heavy connection is necessary for effective transmission, while the connection should be as light as possible when it is intended to serve as a vibration isolator.

Although the results obtained from the longitudinal and lateral modes of vibration are quite similar, there are some basic differences:

(1) In the lateral mode, the relative amplitude also depends on the rotational constraints present at both extremities. In the particular case considered here, the couple at the output end is zero, while the couple at the input end is sufficient to maintain zero slope.

(2) The parameter k depends not only on physical properties but on the distribution of material as well. In fact, k is equal to the product of the quantity c and the radius of gyration.

(3) In both the longitudinal and lateral modes of vibration, the operating frequency range depends on the ratio of the frequency parameter to the length of the connection. The frequency is proportional to this ratio in the former and the square of this quantity in the latter. In both instances, the excitation frequency is directly proportional to the square root of the ratio of the elastic modulus to the density of the material.

The rigidity of the connection (or its converse, the flexibility) may be defined in a relative sense on the basis of the characteristic curves displayed in Figs. 3 and 4. The connection is considered to be "rigid" if the relative amplitude is comparatively uniform, i.e., the relative amplitude is essentially unity over an extended low frequency range, and the increase in successive amplitude peaks is relatively gradual. Conversely, the connection is considered to be "flexible" if the relative amplitude is generally high, i.e., the relative amplitude reaches a high level at a low frequency, and the increase in successive amplitude peaks is quite pronounced. The curves clearly indicate that the rigidity of the connection increases with a decrease in p or q and an increase in s.

It is often possible to maintain a relatively uniform vibration throughout the connection in a particular application or, conversely, to attenuate the vibration level at the load over the entire operating frequency range. This suggests a qualified definition of rigidity and its counterpart, flexibility. Accordingly, we might narrow the concept of transmission to the operating range in which the connection does not experience any phase reversal. Consequently, the connection would be regarded as a transmitter of vibration only when operated in the initial positive segment of the relative amplitude curve, with the limiting condition defined by the existence of a nodal point at the input end. This would serve as a reference boundary, and the suitability of the connection as a transmitter would be enhanced by any change which tended to shift the frequency parameter further away from this arbitrary reference. In a similar fashion, the concept of isolation might be restricted on some occasions to an interval in which the connection does not amplify the input level at any frequency. We accordingly consider that part of the subsequent negative segment of the relative amplitude curve in which the magnitude of the relative amplitude is greater than or equal to negative unity. The efficiency of the connection as an isolator will improve with an increase in the magnitude of the relative amplitude.

The performance characteristics corresponding to the more limited conception of a transmitter and isolator are displayed graphically in Figs. 5 and 6, respectively. In these graphs, the frequency parameter is plotted as a function of the weight ratio for several values of the relative amplitude, which is indicative of the uniformity of a transmitter and the efficiency of an isolator. In both figures, (*a*) applies to the longitudinal and (*b*) to the lateral mode. The graphs indicate that the permissible frequency range diminishes when performance demands become

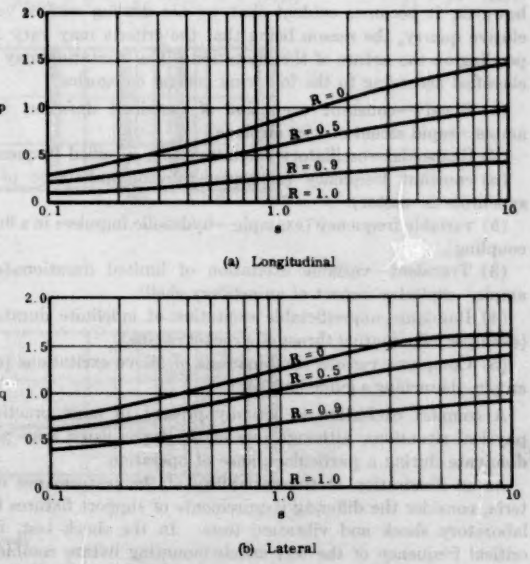
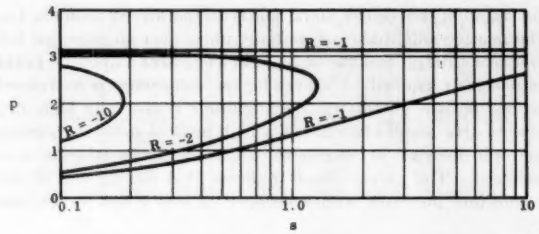
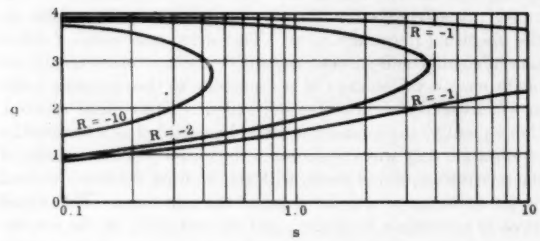


Fig. 5 Performance characteristics of vibration transmitter



(a) Longitudinal



(b) Lateral

Fig. 6 Performance characteristics of vibration isolator

more exacting. For a transmitter, the operable frequency range increases with an increase in the weight ratio, while the opposite is true for an isolator. For a given frequency paramater, an increase in the corresponding system parameter or a reduction in the length of the connection will increase the operating frequency range for both a transmitter and an isolator. However, this extension of the isolation interval will be accompanied by a corresponding elevation of the lowest frequency which can be effectively isolated, and it is this latter consideration which often governs.

Performance Criteria: It is clear that the design of a mechanical connection depends on the objective in mind; specifically whether we wish to transmit the vibration from the source to the load (as in a laboratory test) or, conversely, to isolate the load from the source (as in the service environment). When an attempt is made to formulate criteria of optimum performance, however, it becomes evident that we are dealing with a very elusive quarry, the reason being that the criteria may vary depending on the nature of the excitation. The excitation may be classified according to the following general categories:

(1) Steady—constant excitation of sustained duration (example—rapid ascent of a jet airplane).

(2) Sinusoidal—oscillatory excitation at a specified frequency

(*a*) constant frequency (example—rotating unbalance of a synchronous motor).

(*b*) variable frequency (example—hydraulic impulses in a fluid coupling).

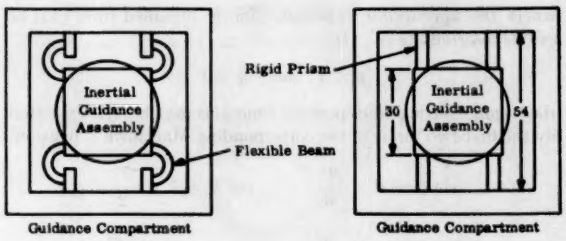
(3) Transient—variable excitation of limited duration (example—explosive impact of an artillery shell).

(4) Random—unpredictable excitation of indefinite duration (example—fluctuating thrust of a rocket engine).

(5) Complex—various combinations of above excitations (example—launching a guided missile).

A complex excitation is actually present in most practical physical situations, although one of the types listed may predominate during a particular phase of operation.

As an illustration of the variability of the performance criteria, consider the differing requirements of support fixtures for laboratory shock and vibration tests. In the shock test, the critical frequency of the test article-mounting fixture combination should be sufficiently high so as not to distort the major part



(a) Soft-Mount (b) Hard-Mount

Fig. 7 Alternative schemes of supporting an inertial guidance assembly in a ballistic missile

of the shock pulse frequency spectrum: Otherwise there may be a substantial discrepancy between the excitation generated by the shock machine and the shock pulse experienced by the test article. In a vibration test, however, which involves a continuous sinusoidal sweep over a broad frequency range, the rigidity of the driver used to transmit the excitation from the shaker to the test article must be sufficient to limit the armature acceleration to an acceptable level. In a random vibration test, the support fixture should be as rigid as possible to expedite the equalization procedure.

Another example is provided by a vibration mount. For a sinusoidal vibration at a constant frequency, it is merely necessary to bracket the excitation frequency in between adjacent critical frequencies, and the isolation efficiency is contingent on the relative amplitude attaining a high value in this interval. In the case of a broad band random excitation, however, the shape of the transmissibility curve over the entire frequency band may assume primary importance, and it is possible that the attenuation provided by an isolator over the major part of the frequency range may be more than offset by the amplification present in a few critical intervals. In this event, it is conceivable that a more rigid support which slightly amplifies the excitation at every frequency might result in a lower output vibration level.

Application of Results

A more concrete illustration of the foregoing comments is provided by the dynamic behavior of a major assembly for an inertial guidance system. The specific problem is to determine a satisfactory means of attaching the assembly to the support rails in the guidance compartment of a ballistic missile. We shall now consider the alternative schemes described schematically in Fig. 7 of employing either a "flexible" beam (*a*) or a "rigid" prism (*b*) for this purpose. A comparison of the corresponding output vibration levels will then enable us to assess the relative merits of a "soft" versus a "hard" mount.

The service excitation consists of a broad band random vibration generated by the rocket engines of the guided missile. We shall assume that the mean square acceleration density [6] is uniform (white noise) over a bandwidth from 10 to 1000 cps. The assembly weighs about 300 lb and is subjected to a "steady" acceleration of 8 g's during launch in addition to fluctuating random inputs.

Soft-Mount. In order to minimize space requirements and cover as broad a frequency range as possible, the fundamental critical frequency of the assembly-isolator system should approach the low end of the excitation frequency band, and 5 cps is accepted as a satisfactory value. The stiffness of the isolator may then be calculated from equation (6) as follows:

$$f_2 = \frac{1}{2\pi}\sqrt{\frac{gK}{W_2}} \quad \text{or} \quad K = \frac{W_2(2\pi f_2)^2}{g}$$

$$K = \frac{(300)[(2)(3.14)(5)]^2}{(386)} = 765 \text{ lb per in.}$$

The isolator is to be made of stainless steel, and any dimensional combination which provides the required stiffness will be satisfactory subject to the restriction that the stress level is not excessive. The maximum bending stress [2] in a cantilever is given by:

$$\sigma = \frac{F_0 L d}{2I} = \frac{3EF_0 d}{2KL^2} \qquad (24)$$

The force (F_0) associated with the static deflection of the assembly is given by:

$$F_0 = W_2 + \frac{W_2}{g}\frac{d^2u_2}{dt^2} = (300) + (300)(8) = 2700 \text{ lb}$$

The corresponding deflection is:

$$\Delta = \frac{F_0}{K} = \frac{(2700)}{(765)} = 3.5 \text{ in.}$$

If the stress level is not to exceed 75,000 psi, the proportions of the cantilever must satisfy equation (24) as follows:

$$\frac{L^2}{d} = \frac{3EF_0}{2K\sigma} = \frac{(3)(28 \times 10^6)(2700)}{(2)(765)(75{,}000)} = 2000$$

while in order to provide the necessary stiffness, equation (4*b*) requires that:

$$\frac{L^3}{bd^3} = \frac{E}{4K} = \frac{(28 \times 10^6)}{(4)(765)} = 9150$$

These conditions are met by a cantilever of the following dimensions:

$$L = 25 \text{ in.} \qquad d = {}^5/_{16} \text{ in.} \qquad b = 56 \text{ in.}$$

The quantities required to determine the relative amplitude as a function of the excitation frequency are computed from equations (13), (16), and (22) as follows:

$$w = \gamma bdL = (0.286)(56)({}^5/_{16})(25) = 125 \text{ lb}$$

$$s = \frac{w}{W_2} = \frac{(125)}{(300)} = 0.415$$

$$K = \frac{d}{2}\sqrt{\frac{gE}{3\gamma}} = \frac{(5)}{(2)(16)}\sqrt{\frac{(386)(28 \times 10^6)}{(3)(0.286)}} = 1.75 \times 10^4$$

$$q = L\sqrt{\frac{2\pi f}{k}} = (25)\sqrt{\frac{(2)(3.14)f}{1.75 \times 10^4}} = 0.475\sqrt{f}$$

Since $q = 1.06 \approx 1$ and $s = 0.415 \ll 2.67$ when $f = 5$ cps, the use of the simplified relation (6) to approximate the lowest critical frequency is justified. The transmissibility is then calculated from equation (21) and the result graphed in Fig. 8(*a*). Although effective isolation is provided over the major part of the excitation frequency range, the input is amplified in a number of relatively narrow frequency intervals.

Hard-Mount. In view of the deficiencies of the soft-mount, the possibility of using a more rigid connection will now be investigated. The hard-mount is assumed to be made of the same material as the soft-mount with the same overall weight. However, the effective length of the connection is now determined by the distance between the assembly and the mounting rails in the guidance compartment rather than stiffness and strength considerations. As a result, the 25-in. long beam is replaced by a 12-in. long prism. The pertinent quantities are calculated from equations (7) and (13) as follows:

$$s = \frac{w}{W_2} = \frac{(125)}{(300)} = 0.415$$

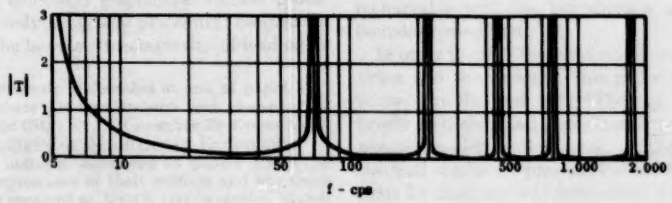
$$c = \sqrt{\frac{gE}{\gamma}} = \sqrt{\frac{(386)(28 \times 10^6)}{(0.286)}} = 1.95 \times 10^6$$

$$p = \frac{2\pi Lf}{c} = \frac{(2)(3.14)(12)f}{(1.95 \times 10^6)} = 3.85 \times 10^{-6}f$$

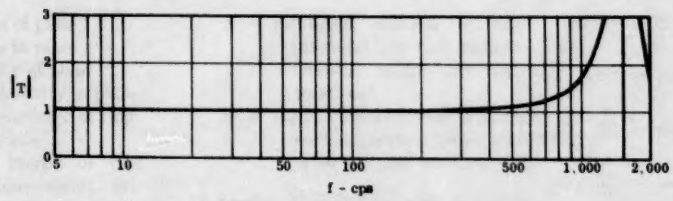
The transmissibility is calculated from equation (15) and the result graphed in Fig. 8(*b*). While the input is now amplified to some extent over the entire excitation frequency range, the sharp peaks have been eliminated.

Comparison of Hard Versus Soft-Mount. The rms (root-mean-square) input and output levels may be computed [6] for a random vibration as follows:

$$\ddot{\lambda}_{1\text{rms}} = g\sqrt{\int_{f_a}^{f_b} G\,df} = g\sqrt{GB} \qquad B = f_b - f_a \qquad (25a)$$



(a) Soft-Mount



(b) Hard-Mount

Fig. 8 Frequency response of inertial guidance assembly

$$\ddot{\lambda}_{2\text{rms}} = g\sqrt{\int_{fa}^{fb} GT^2 df} = g\sqrt{GJ} \qquad J = \int_{fa}^{fb} T^2 df \quad (25b)$$

where G is the mean square acceleration density of the random excitation. A modified rms transmissibility may then be defined as follows:

$$Q_{\text{rms}} = \frac{\ddot{\lambda}_{2\text{rms}}}{\ddot{\lambda}_{1\text{rms}}} = \sqrt{\frac{J}{B}} = T_{\text{rms}} \quad (26)$$

which reveals that the rms transmissibility is equal to the rms value of the transmissibility curve for a "white noise" random excitation. An estimate of the quantity for the soft and hard-mount discloses that the former is superior to the latter in this particular application provided natural damping present in the system is sufficient to limit the peak transmissibility to moderate values ($T < 40$). However, it is interesting to find that the advantage enjoyed by the soft-mount is rather modest. Perhaps the most noteworthy difference between the hard and soft-mount is due to the filtering action of the latter. This tends to concentrate the output at the critical frequencies of the assembly-isolator system, whereas the hard-mount uniformly distributes the output over the entire frequency range.

Conclusions

The results of the preceding analysis may be briefly summarized as follows:

(1) In a literal sense, a mechanical connection behaves in a rigid manner only at excitation frequencies which are low compared with the critical frequency of the load-connection combination. Isolation occurs in frequency intervals separated by gaps in which the vibration is amplified.

(2) It is sometimes necessary to operate a system in the isolation range when a rigid connection is desired, and also to include intervals of undesirable amplification when isolation is required. Consequently, a more generalized conception of rigidity based on the characteristic behavior of the system is appropriate.

(3) The essential difference between a simplified analysis which ignores the distributed mass of the connection and a more refined treatment of the problem consists of an alteration of the response of the system from a monotonic to an oscillatory characteristic. The results of the simplified analysis are a good approximation of the actual behavior of the system only for relatively moderate values of the frequency parameter and weight ratio.

(4) The suitability of a connection as a vibration transmitter improves with an increase in the weight ratio and a decrease in the operating frequency. Its effectiveness as a vibration isolator improves with a decrease in the weight ratio and a reduction in the isolation range. The latter is accompanied by an elevation of the lowest (and a corresponding depression of the highest) frequency which can be effectively isolated.

(5) It is difficult to formulate general performance criteria for an effective vibration transmitter or isolator because of the variability of the criteria with the nature of the excitation. In an illustrative example, a soft-mount was found to be superior to a hard-mount as a vibration isolator.

Acknowledgments

The co-operation and assistance of various personnel of AC Spark Plug Division of General Motors Corporation is gratefully acknowledged.

References

1 W. T. Thomson, "Mechanical Vibrations," Prentice-Hall Company, Englewood Cliffe, N. J., second edition, 1956, p. 252.

2 R. J. Roark, "Formulas for Stress and Strain," McGraw-Hill Book Company, Inc., New York, N. Y., third edition, 1954, p. 381.

3 C. E. Crede, "Vibration and Shock Isolation," John Wiley & Sons, Inc., New York, N. Y., 1957, p. 328.

4 S. Timoshenko and D. H. Young, "Vibration Problems in Engineering," D. Van Nostrand and Company, New York, N. Y., third edition, 1955, p. 468.

5 K. S. Miller, "Partial Differential Equations in Engineering Problems," Prentice-Hall Company, New York, N. Y., 1953, p. 254.

6 "Random Vibration," edited by S. H. Crandall, The Technology Press, M.I.T., Cambridge, Mass., 1958.

Deformation and Moments in Elastically Restrained Circular Plates Under Arbitrary Load or Linear Thermal Gradient

M. ZAID
Scientific Research Staff
Assoc. Mem. ASME

M. FORRAY
Structural Methods Staff

Republic Aviation Corporation,
Farmingdale, N. Y.

The problem of a circular plate, with a central hole, elastically constrained against rotation and deflection, acted upon by a transverse linear temperature gradient or a general axisymmetrical loading condition is considered in this paper. With the aid of presented graphs and simple algebraic equations, it is a relatively simple matter to construct the desired deflection and stress pattern for any combination of elastic constraints and load distribution.

INTRODUCTION

THE DESIGNER is frequently called upon to select a thin circular plate which is acted upon by a transverse linear temperature gradient or an axisymmetric loading condition. In the event that the loading is simple and the plate is either simply supported or clamped, equations are readily available for the evaluation of stress and deflection [1].[1] However, in the more realistic cases, the supports are never simple or clamped and the loading is often complex, Fig. 1. The designer must then decide how to approximate the actual structure so that the elementary results can be used. The question of whether the plate is simply supported or clamped, or whether the loading is essentially uniform, becomes quite difficult to answer without the experience acquired by experimental verification of many previous solutions. Alternatively, the designer might be asked to select a minimum weight support that leads to stress and deformation results that are within a specified percentage of the theoretical simple or clamped condition.

It is the purpose of this paper to present graphs which, when combined with simple algebraic formulas, will enable the designer to rapidly obtain the stress and deformation throughout the plate for any axisymmetric transverse loading or linear temperature gradient, and any combination of elastic supports.

The techniques of this combined graphical-analytical procedure (as compared to a purely analytical procedure) become increasingly valuable with the increase in complexity of loading or boundary condition. This results from the fact that the time and effort required to obtain a complete stress and deformation solution by the techniques of this paper is essentially independent of the complexity of the problem. For example, once some facility is obtained, the complete solution to the most complex problem could easily be completed in less than two days. On the other hand, the purely analytical approach becomes quite unmanageable with the introduction of any complex loading or boundary condition.

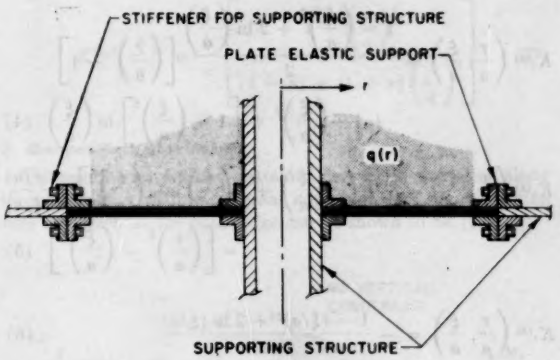


Fig. 1

In order to make the ideas readily accessible to both the theoretician and the designer, this paper has been divided into two parts. In the first part (Theory) fundamental solutions are briefly indicated and then, based on these, the solution to the general problem is developed. In the second part (Applications) the final results are presented in the form of graphs with instructions for their use, and in addition a typical problem is solved to illustrate the application of these results.

[1] Numbers in brackets designate References at end of paper.

Contributed by the Machine Design Division and presented at the Annual Meeting, Atlantic City, N. J., November 29–December 4, 1959, of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS.

NOTE: Statements and opinions advanced in papers are to be understood as individual expressions of their authors and not those of the Society. Manuscript received at ASME Headquarters, March 18, 1959. Paper No. 59—A-39.

Nomenclature

a = outside radius of plate
b = radius of hole in plate
D = flexural rigidity of plate
E = modulus of elasticity of plate
E_r = modulus of elasticity of ring
h = thickness of plate
I_{zz} = moment of inertia of ring cross section about axis through c.g., perpendicular to the axial direction of the ring
k = rotational stiffness of ring (moment per unit circumferential length per unit rotation)[2]
K = translational stiffness of members supporting plate edge (load per unit deflection)[2,3]

[2] Subscript "i" indicates inside, subscript "o" indicates outside.

[3] Starred quantity indicates value for reduced system.

$K^{(0)}$ = $K^{(i)}$ associated with the total line load P_j at ξ_j
$K^{(Q)}$ = $K^{(i)}$ associated with the redundant line load (Q)
M_1, M_2 = redundant moments per unit circumferential length
M_r, M_t = radial and tangential moment per unit circumferential length

(*Continued on next page*)

THEORY

A Elementary Solutions

1 Externally Clamped Plate, General Loading

Using the known plate equation [1], it can be shown that the solution for the solid clamped plate acted upon by the incremental total line load $dP = 2\pi q\xi d\xi$, Fig. 2, is given by

$$\frac{4D}{a^2}\,dw\left(\frac{r}{a}\right) = \xi q\left(\frac{\xi}{a}\right)d\xi K_j^{(1)}\left(\frac{r}{a},\frac{\xi}{a}\right)_{(j=1,2)} \qquad 0 \le \frac{r}{a} \le 1 \tag{1}$$

$$2\,\frac{dM_r}{1+\nu} = \xi q\left(\frac{\xi}{a}\right)d\xi K_j^{(2)}\left(\frac{r}{a},\frac{\xi}{a}\right)_{(j=1,2)} \qquad 0 \le \frac{r}{a} \le 1 \tag{2}$$

$$2\,\frac{dM_t}{1+\nu} = \xi q\left(\frac{\xi}{a}\right)d\xi K_j^{(3)}\left(\frac{r}{a},\frac{\xi}{a}\right)_{(j=1,2)} \qquad 0 \le \frac{r}{a} \le 1 \tag{3}$$

where

$$K_1^{(1)}\left(\frac{r}{a},\frac{\xi}{a}\right) = \frac{1-\left(\frac{\xi}{a}\right)^2+2\ln\left(\frac{\xi}{a}\right)}{2}\left[\left(\frac{r}{a}\right)^2-1\right] + 1-\left(\frac{\xi}{a}\right)^2+\left[1+\left(\frac{\xi}{a}\right)^2\right]\ln\left(\frac{\xi}{a}\right) \tag{4}$$

$$K_2^{(1)}\left(\frac{r}{a},\frac{\xi}{a}\right) = K_1^{(1)}\left(\frac{r}{a},\frac{\xi}{a}\right)+\left[\left(\frac{r}{a}\right)^2+\left(\frac{\xi}{a}\right)^2\right]\ln\frac{(r/a)}{(\xi/a)} - \left[\left(\frac{r}{a}\right)^2-\left(\frac{\xi}{a}\right)^2\right] \tag{5}$$

$$K_1^{(2)}\left(\frac{r}{a},\frac{\xi}{a}\right) = -\frac{1-(\xi/a)^2+2\ln(\xi/a)}{2} \tag{6}$$

$$K_2^{(2)}\left(\frac{r}{a},\frac{\xi}{a}\right) = K_1^{(2)}-\frac{1}{2}\left[\frac{(1-\nu)}{(1+\nu)}\left(1-\frac{(\xi/a)^2}{(r/a)^2}\right)+2\ln\frac{(r/a)}{(\xi/a)}\right] \tag{7}$$

$$K_1^{(3)}\left(\frac{r}{a},\frac{\xi}{a}\right) = -\frac{1-(\xi/a)^2+2\ln(\xi/a)}{2} \tag{8}$$

$$K_2^{(3)}\left(\frac{r}{a},\frac{\xi}{a}\right) = K_1^{(3)}+\frac{1}{2}\left[\frac{1-\nu}{1+\nu}\left(1-\frac{(\xi/a)^2}{(r/a)^2}\right)-2\ln\frac{(r/a)}{(\xi/a)}\right] \tag{9}$$

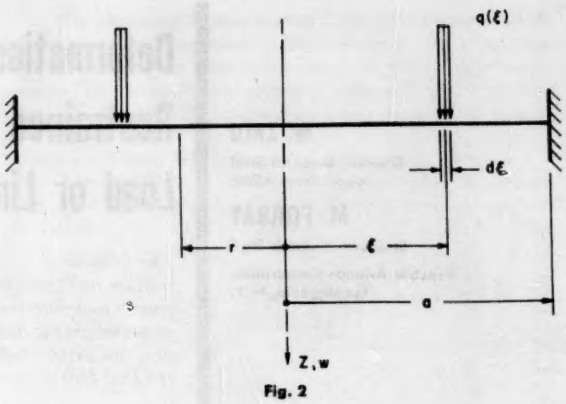


Fig. 2

and the region of applicability is,

$$K_1^{(i)};\quad 0 \le \frac{r}{a} \le \frac{\xi}{a} \le 1 \tag{10}$$

$$K_2^{(i)};\quad 0 \le \frac{\xi}{a} \le \frac{r}{a} \le 1 \tag{11}$$

These values of $K_j^{(i)}$ are plotted at the end of the paper and are labeled without the subscript "j".

Using the reciprocal theorem of Betti [2], or from direct calculation in equations (4) and (5), it follows that

$$K_1^{(1)}\left(\frac{r}{a},\frac{\xi}{a}\right) = K_2^{(1)}\left(\frac{\xi}{a},\frac{r}{a}\right) \tag{12}$$

$$K_2^{(1)}\left(\frac{r}{a},\frac{\xi}{a}\right) = K_1^{(1)}\left(\frac{\xi}{a},\frac{r}{a}\right) \tag{13}$$

Equations (12) and (13) are useful in interpolating and drawing additional $K_j^{(1)}$ curves, if required, for any special problem.

For future work it is necessary to have the deflection and moment at any point in a solid plate due to a load which is arbitrary for $r/a \ge b/a$ and which is zero for $r/a < b/a$. This is obtained from superposition, using equations (1) to (3) inclusive, so that

$$\frac{8\pi D}{P_1 a^2}\,w\left(\frac{r}{a}\right) = \int_{b/a}^{1}\left(\frac{\xi}{a}\right)\frac{q\left(\frac{\xi}{a}\right)}{q_1}K_j^{(1)}\left(\frac{r}{a},\frac{\xi}{a}\right)d\left(\frac{\xi}{a}\right) \equiv K^{(1)1}\left(\frac{r}{a}\right) \tag{14}$$

Nomenclature

P = total external force due to distributed load only, taken positive in positive z-direction

P_{total} = total external force on unreduced system, taken positive in positive z-direction

$P_1 = 2\pi a^2 q_1$

P_j = total line load at position ξ_j, taken positive in positive z-direction

$q(\xi/a)$ = external distributed load per unit area, taken positive in positive z-direction

q_1 = arbitrary assigned distributed load per unit area, taken positive in positive z-direction

Q = redundant force per unit circumferential length

Q_{total} = total force per unit circumferential length, acting on a plate at a generic radius

$\bar{r}$ = radius of ring (measured to c.g.)

r = radial co-ordinate

t = steady-state temperature gradient $\left[t\left(\frac{h}{2}\right)-t\left(-\frac{h}{2}\right)\right]$

$t\left(\frac{h}{2}\right), t\left(-\frac{h}{2}\right)$ = temperature at the plate surfaces

w = deflection[2]

α = coefficient of linear expansion

θ = rotation of ring about centroid of cross section

ξ = radial co-ordinate which defines position of load

σ_r, σ_t = radial and tangential stress

τ = shear stress

ν = Poisson's ratio

$$\frac{4\pi M_r}{P_1(1+\nu)} = \int_{b/a}^{1} \left(\frac{\xi}{a}\right) \frac{q\left(\frac{\xi}{a}\right)}{q_1} K_j^{(3)}\left(\frac{r}{a}, \frac{\xi}{a}\right) d\left(\frac{\xi}{a}\right) \equiv K^{(3)1}\left(\frac{r}{a}\right) \quad (15)$$

$$\frac{4\pi M_t}{P_1(1+\nu)} = \int_{b/a}^{1} \left(\frac{\xi}{a}\right) \frac{q\left(\frac{\xi}{a}\right)}{q_1} K_j^{(3)}\left(\frac{r}{a}, \frac{\xi}{a}\right) d\left(\frac{\xi}{a}\right) \equiv K^{(3)1}\left(\frac{r}{a}\right) \quad (16)$$

where q_1 is arbitrarily assigned and $P_1 = 2\pi a^2 q_1$. It is sometimes convenient to take $q_1 = q_{\max}$ so that $0 \leq \frac{q}{q_1} \leq 1$. As a special useful case the results for a finite line load at ξ_j, having a total force given by $P_j = \lim_{\epsilon \to 0} 2\pi \int_{\xi_j - \epsilon}^{\xi_j + \epsilon} \xi q(\xi) d\xi$ are

$$\frac{8\pi D}{P_j a^2} w\left(\frac{r}{a}\right) = K_j^{(1)}\left(\frac{r}{a}, \frac{\xi_j}{a}\right) \equiv K^{(1)j}\left(\frac{r}{a}\right) \quad (17)$$

$$\frac{4\pi M_r\left(\frac{r}{a}\right)}{P_j(1+\nu)} = K_j^{(2)}\left(\frac{r}{a}, \frac{\xi_j}{a}\right) \equiv K^{(2)j}\left(\frac{r}{a}\right) \quad (18)$$

$$\frac{4\pi M_t\left(\frac{r}{a}\right)}{P_j(1+\nu)} = K_j^{(3)}\left(\frac{r}{a}, \frac{\xi_j}{a}\right) \equiv K^{(3)j}\left(\frac{r}{a}\right) \quad (19)$$

$$(j = 2, 3, \ldots n)$$

For future work it is necessary to have the slope at $r/a = b/a$ for a clamped plate with a central hole. This can be expressed as

$$\frac{4\pi D}{P_1 a}\left(\frac{dw}{dr}\right) = \int_{b/a}^{1} \left(\frac{\xi}{a}\right) \frac{q\left(\frac{\xi}{a}\right)}{q_1} K^{(4)}\left(\frac{\xi}{a}\right) d\left(\frac{\xi}{a}\right) \equiv K^{(4)1}\left(\frac{b}{a}\right) \quad (20)$$

where

$$K^{(4)} = \frac{b}{a}\left[\frac{1 - \left(\frac{\xi}{a}\right)^2 + 2\ln\left(\frac{\xi}{a}\right)}{(1-\nu) + (1+\nu)\left(\frac{b}{a}\right)^2}\right] \quad (21)$$

and for the special case of a line load at ξ_j

$$\frac{4\pi D}{P_j a}\left(\frac{dw}{dr}\right)_{\frac{r}{a}=\frac{b}{a}} = K^{(4)}\left(\frac{\xi_j}{a}\right) \equiv K^{(4)j}\left(\frac{b}{a}\right) \quad (22)$$

In particular, where $b \to 0$, the proper evaluation of the limit in $K^{(4)}$ leads to $\lim_{b \to 0^+}\left[\frac{dw}{dr}\right] = 0$ independent of the loading.

The determination of the deformations and moments for an arbitrary load, using the graphs for $K^{(4)}$ and equations (14) to (16) and (20) will be discussed in the section Applications.

2 Clamped Plate, Central Hole, Thermal Gradient

It can be shown that the solution for the solid, externally clamped, plate acted upon by a transverse thermal gradient is given by

$$w\left(\frac{r}{a}\right) = 0 \quad (23)$$

$$M_r\left(\frac{r}{a}\right) = -\frac{D\alpha t}{h}(1+\nu) \quad (24)$$

$$M_t\left(\frac{r}{a}\right) = -\frac{D\alpha t}{h}(1+\nu) \quad (25)$$

For future work it is necessary to have the slope at $r/a = b/a$ for the plate with an unrestrained central hole. This can be expressed as

$$\frac{h}{a\alpha t(1+\nu)}\left(\frac{dw}{dr}\right)_{\frac{r}{a}=\frac{b}{a}} = K^{(5)}\left(\frac{b}{a}\right) \quad (26)$$

where

$$K^{(5)}\left(\frac{b}{a}\right) = \frac{b}{a}\left[\frac{1 - \left(\frac{b}{a}\right)^2}{(1-\nu) + (1+\nu)\left(\frac{b}{a}\right)^2}\right] \quad (27)$$

3 Clamped Plate, Edge Moment

The solution for a plate, restrained only against rotation along its inner radius, with an external edge moment, Fig. 3, having zero deflection at the outer edge can be shown to be [1],

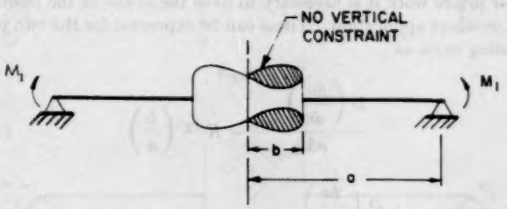


Fig. 3

$$\frac{2Dw}{a^2 M_1} = K^{(6)}\left(\frac{r}{a}\right) \quad (28)$$

$$\frac{M_r}{M_1} = K^{(7)}\left(\frac{r}{a}\right) \quad (29)$$

$$\frac{M_t}{M_1} = K^{(8)}\left(\frac{r}{a}\right) \quad (30)$$

where

$$K^{(6)}\left(\frac{r}{a}\right) = \frac{1 - (r/a)^2 + 2(b/a)^2 \ln(r/a)}{1 + \nu + (b/a)^2(1-\nu)} \quad (31)$$

$$K^{(7)}\left(\frac{r}{a}\right) = \frac{1 + \nu + \frac{(b/a)^2}{(r/a)^2}(1-\nu)}{1 + \nu + (b/a)^2(1-\nu)} \quad (32)$$

$$K^{(8)}\left(\frac{r}{a}\right) = \frac{1 + \nu - \frac{(b/a)^2}{(r/a)^2}(1-\nu)}{1 + \nu + (b/a)^2(1-\nu)} \quad (33)$$

The solution for the externally clamped plate with only an internal edge moment, Fig. 4, can be shown to be [1],

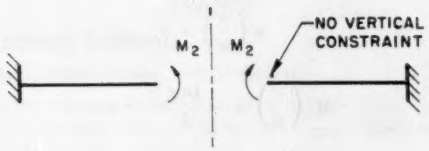


Fig. 4

$$\frac{2Dw}{a^2M_2} = K^{(9)}\left(\frac{r}{a}\right) \tag{34}$$

$$\frac{M_r}{M_2} = K^{(10)}\left(\frac{r}{a}\right) \tag{35}$$

$$\frac{M_t}{M_2} = K^{(11)}\left(\frac{r}{a}\right) \tag{36}$$

where

$$K^{(9)} = \frac{(b/a)^2[1 - (r/a)^2 + 2\ln(r/a)]}{1 - \nu + (1+\nu)(b/a)^2} = \frac{b}{a}K^{(4)}\left(\frac{r}{a}\right) \tag{37}$$

$$K^{(10)} = \frac{(b/a)^2}{(r/a)^2}\frac{[(1-\nu) + (1+\nu)(r/a)^2]}{[(1-\nu) + (1+\nu)(b/a)^2]} \tag{38}$$

$$K^{(11)} = -\frac{(b/a)^2}{(r/a)^2}\frac{[(1-\nu) - (1+\nu)(r/a)^2]}{[(1-\nu) + (1+\nu)(b/a)^2]} \tag{39}$$

For future work it is necessary to have the slopes at the position of moment application. These can be expressed for the two preceding cases as

$$\frac{D\left(\frac{dw}{dr}\right)_{r/a=1}}{aM_1} = K^{(12)}\left(\frac{b}{a}\right) \tag{40}$$

$$\frac{D\left(\frac{dw}{dr}\right)_{r/a=b/a}}{aM_2} = K^{(13)}\left(\frac{b}{a}\right) \tag{41}$$

where

$$K^{(12)}\left(\frac{b}{a}\right) = -\frac{1 - (b/a)^2}{(1+\nu) + (1-\nu)(b/a)^2} \tag{42}$$

$$K^{(13)}\left(\frac{b}{a}\right) = \frac{(b/a)[1-(b/a)^2]}{(1-\nu) + (1+\nu)(b/a)^2} = K^{(5)}\left(\frac{b}{a}\right) \tag{43}$$

The values of $K^{(i)}$ are plotted at the end of the paper.

4 Thin Ring, Edge Moment

Considering small deformations, the rotation of a ring acted upon by a constant circumferential moment, Fig. 5, is in the direction of and proportional to that moment. This is given by

$$M = k\theta \tag{44}$$

Fig. 5

where k is the unit torsional stiffness of the system and M is the moment per unit circumference.

The value of k can be determined either experimentally or analytically. A good approximate value of k for rings having a thickness which is small compared to the radius to the center of gravity is [3],

$$k = \frac{E_R I_{zz}}{\bar{r}^2} \tag{45}$$

B General Solution

The general solution for any pressure loading or thermal gradient associated with a combination of elastic boundary conditions can be obtained from a suitable superposition of the elementary solutions presented in the foregoing. This superposition must be such as to satisfy all the necessary boundary conditions on deflection, slope, shear force, and moment.

1 Statement of Most General Problem

The general problem with two arbitrary rotational constraints and two arbitrary translational constraints is shown in Fig. 6.

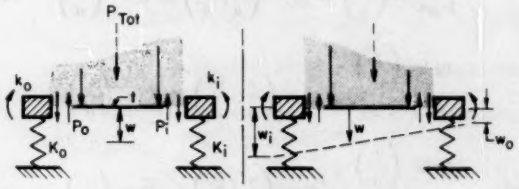


Fig. 6 General problem

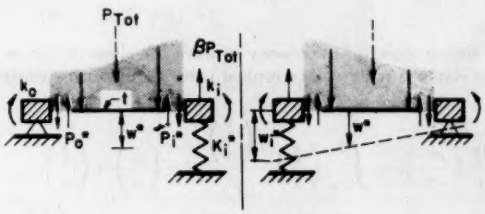


Fig. 7 Reduced system

It will be shown in this section that the plate associated with the reduced system, Fig. 7, when compared to that in the general system, Fig. 6, has an identical stress distribution, and a displacement pattern that differs only by an additive constant. Thus the system with two translational constraints can be reduced to a system with a fixed outer support, an inner support with modified rigidity (K_i^*), and an edge force (βP_{tot}) on the inner circumference; all other loadings and constraints remain the same. It should be carefully noted that, in this paper, the force (βP_{tot}) is used in conjunction with the boundary condition, and is not taken as an additional plate load.

Proof:

Let us assume that w is the solution to the general system and that the solution for the reduced system is given by,

$$w^* = w - w_o \tag{46}$$

then it follows that, since $\frac{d^n w}{dr^n} = \frac{d^n w^*}{dr^n}$, the basic plate differential equation and all rotational boundary conditions are identically satisfied for w^*. From this it follows that the stresses are identical in both plates. It is now only necessary to show that w^* is compatible with the displacement boundary constraints and in addition does not violate the force boundary conditions. Be-

cause of the identical stress distribution these force boundary conditions must be identical for both the reduced and unreduced system.

The only condition on displacement in the reduced system is that $w^* = 0$ on the outer edge. Since $w = w_o$ at the outer edge, the use of equation (46) leads to the satisfaction of the displacement condition. It remains only to determine the elastic constraint K_i^* and the additional edge force βP_{tot} which will lead to identical total edge forces. Defining, as positive, the downward (positive z) external forces and, in addition, the resulting upward resisting force on the plate, the relevant equations for the unreduced system are:

Spring equations:

$$P_i = K_i w_i \tag{47}$$

$$P_o = K_o w_o \tag{48}$$

where P_i and P_o are the total upward edge forces on the plate.

The similar spring equations for the reduced system are:

$$P_i^* = \beta P_{tot} + K_i^* w_i^* \tag{49}$$

$$P_o^* = P_{tot} - P_i^* \tag{50}$$

where P_{tot} is the total downward force acting on the original system. In addition, from equation (46) there results

$$w_i^* = w_i - w_o \tag{51}$$

The force boundary conditions are

$$P_o^* = P_o \tag{52}$$

$$P_i^* = P_i \tag{53}$$

Reducing this system of equations to one involving the parameters w_i, P_{tot}, β, and K_i^*, and considering that w_i can be varied independently of P_{tot} (i.e., it is a function of the distribution of q as well as of the temperature gradient and boundary conditions), there results

$$\beta = \frac{1}{1 + \dfrac{K_o}{K_i}} \tag{54}$$

$$K_i^* = K_o \frac{1}{1 + \dfrac{K_o}{K_i}} \tag{55}$$

and in addition,

$$w_o = \beta \frac{P_{tot}}{K_i} - \beta w_i^* \tag{56}$$

The deflection in the actual plate can be evaluated from equation (46) once the value for w^* is obtained for the reduced problem. In all further discussion the reduced problem will be considered as the general problem.

2 Superposition Equations

The solution to the general problem, Fig. 7, can be obtained by the superposition of the problem types shown in Figs. 8(a, b, c, and d). In doing this it is convenient to solve the problem in Fig. 8(a) by the superposition of the two simpler problem types shown in Figs. 8(a' and a''). The unknown quantities Q, M_1, and M_2 are evaluated in such a manner as to satisfy the linear force or moment—deflection or slope relationship associated with the elastic constraint.

From equilibrium consideration these are

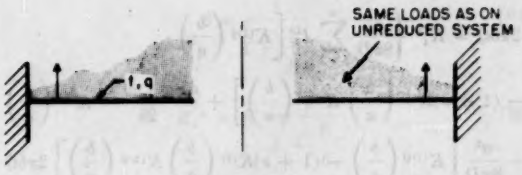


Fig. 8(a)

Fig. 8(a')

Fig. 8(a'')

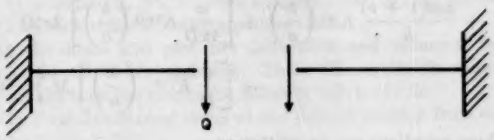


Fig. 8(b)

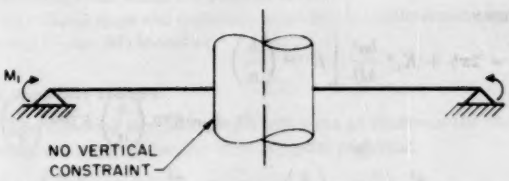


Fig. 8(c)

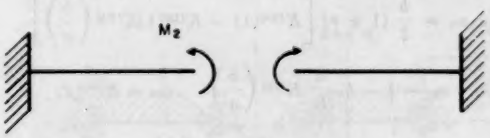


Fig. 8(d)

$$P_i^* = -2\pi b Q = K_i^* w_{itot}^* + \beta P_{tot} \tag{57}$$

$$M_{otot} = k_o \left(\frac{dw^*}{dr}\right)_{otot} \tag{58}$$

$$M_{itot} = -k_i \left(\frac{dw^*}{dr}\right)_{itot} \tag{59}$$

The condition of zero deflection at the outside support is automatically satisfied. Using the previous solutions, and expressing the deflection and slopes in terms of the previously evaluated $K^{(0)}$ and the unknown forces and moments, equations (57) to (59) inclusive become,

$$-2\pi bQ = K_i^* \left\{ \frac{a^2}{8\pi D} \sum_{j=1}^{n} P_j \left[K^{(1)j}\left(\frac{b}{a}\right) - (1+\nu) K^{(8)}\left(\frac{b}{a}\right) K^{(2)j}\left(\frac{b}{a}\right) \right] + \frac{a^2 \alpha t (1+\nu)}{2h} K^{(9)}\left(\frac{b}{a}\right) + \frac{a^2}{8\pi D}\left[K^{(1)Q}\left(\frac{b}{a}\right) - (1+\nu) K^{(8)}\left(\frac{b}{a}\right) K^{(2)Q}\left(\frac{b}{a}\right) \right] 2\pi bQ + \left[\frac{a^2}{2D} K^{(8)}\left(\frac{b}{a}\right) \right] M_1 + \left[\frac{a^2}{2D} K^{(9)}\left(\frac{b}{a}\right) \right] M_2 \right\} + \beta P_{tot} \tag{60}$$

$$\frac{1+\nu}{4\pi} \sum_{j=1}^{n} P_j \left[K^{(2)j}(1) - K^{(10)}(1) K^{(2)j}\left(\frac{b}{a}\right) \right] + \frac{D\alpha t(1+\nu)}{h} [K^{(10)}(1) - 1] + \frac{(1+\nu)}{4\pi} \left[K^{(2)Q}(1) - K^{(10)}(1) K^{(2)Q}\left(\frac{b}{a}\right) \right] 2\pi bQ + M_1 + K^{(10)}(1) M_2 = k_o \left[\frac{a}{D} K^{(12)}\left(\frac{b}{a}\right) \right] M_1 \tag{61}$$

$$\left[K^{(7)}\left(\frac{b}{a}\right) \right] M_1 + M_2 = -k_i \left\{ \frac{a}{4\pi D} \sum_{j=1}^{n} P_j K^{(4)j}\left(\frac{b}{a}\right) + \frac{a\alpha t(1+\nu)}{h} K^{(5)}\left(\frac{b}{a}\right) + \left[\frac{a}{4\pi D} K^{(4)Q}\left(\frac{b}{a}\right) \right] 2\pi bQ + \left[\frac{a}{D} K^{(13)}\left(\frac{b}{a}\right) \right] M_2 \right\} \tag{62}$$

These equations can be written as

$$a_{ij} V_j = b_i \tag{63}$$

where

$$a_{11} = 2\pi b + K_i^* \frac{ba^2}{4D} \left[K^{(1)Q}\left(\frac{b}{a}\right) - (1+\nu) K^{(8)}\left(\frac{b}{a}\right) K^{(2)Q}\left(\frac{b}{a}\right) \right]$$

$$a_{12} = \frac{a^2}{2D} K_i^* K^{(6)}\left(\frac{b}{a}\right) \qquad a_{13} = \frac{a^2}{2D} K_i^* K^{(9)}\left(\frac{b}{a}\right)$$

$$a_{21} = \frac{b}{2}(1+\nu)\left[K^{(2)Q}(1) - K^{(10)}(1) K^{(2)Q}\left(\frac{b}{a}\right) \right]$$

$$a_{22} = 1 - k_o \frac{a}{D} K^{(12)}\left(\frac{b}{a}\right) \qquad a_{23} = K^{(10)}(1)$$

$$a_{31} = k_i \frac{ab}{2D} K^{(4)Q}\left(\frac{b}{a}\right)$$

$$a_{32} = K^{(7)}\left(\frac{b}{a}\right) \qquad a_{33} = 1 + k_i \frac{a}{D} K^{(13)}\left(\frac{b}{a}\right)$$

$$V_1 = Q \qquad V_2 = M_1 \qquad V_3 = M_2$$

$$b_1 = -K_i^* \frac{a^2}{8\pi D} \sum_{j=1}^{n} P_j \left[K^{(1)j}\left(\frac{b}{a}\right) - (1+\nu) K^{(8)}\left(\frac{b}{a}\right) K^{(2)j}\left(\frac{b}{a}\right) \right] - K_i^* \frac{a^2 \alpha t(1+\nu)}{2h} K^{(9)}\left(\frac{b}{a}\right) - \beta P_{tot}$$

$$b_2 = -\frac{1+\nu}{4\pi} \sum_{j=1}^{n} P_j \left[K^{(2)j}(1) - K^{(10)}(1) K^{(2)j}\left(\frac{b}{a}\right) \right] + \frac{D\alpha t(1+\nu)}{h} [1 - K^{(10)}(1)]$$

$$b_3 = -k_i \left[\frac{a}{4\pi D} \sum_{j=1}^{n} P_j K^{(4)j}\left(\frac{b}{a}\right) + \frac{a\alpha t(1+\nu)}{h} K^{(5)}\left(\frac{b}{a}\right) \right] \tag{64}$$

where all summations are over the loads in the unreduced system only. The solution for the redundants can be expressed in terms of these coefficients as,

$$Q = \frac{b_1(a_{22}a_{33} - a_{23}a_{32}) + b_2(a_{13}a_{32} - a_{12}a_{33}) + b_3(a_{12}a_{23} - a_{13}a_{22})}{\Delta} \tag{65}$$

$$M_1 = \frac{b_1(a_{23}a_{31} - a_{21}a_{33}) + b_2(a_{11}a_{33} - a_{13}a_{31}) + b_3(a_{13}a_{21} - a_{11}a_{23})}{\Delta} \tag{66}$$

$$M_2 = \frac{b_1(a_{21}a_{32} - a_{22}a_{31}) + b_2(a_{12}a_{31} - a_{11}a_{32}) + b_3(a_{11}a_{22} - a_{12}a_{21})}{\Delta} \tag{67}$$

where

$$\Delta = a_{11}(a_{22}a_{33} - a_{23}a_{32}) + a_{21}(a_{13}a_{32} - a_{12}a_{33}) + a_{31}(a_{12}a_{23} - a_{13}a_{22}) \tag{68}$$

Once the values of these redundants are determined, the total deflections and moments at any position (r/a) can be obtained through the superposition of the previously indicated elementary solutions.

Thus

$$w_{tot}^* = \frac{a^2}{8\pi D} \sum_{j=1}^{n} P_j \left[K^{(1)j}\left(\frac{r}{a}\right) - (1+\nu) K^{(9)}\left(\frac{r}{a}\right) K^{(2)j}\left(\frac{b}{a}\right) \right] + \frac{a^2 b}{4D} Q \left[K^{(1)Q}\left(\frac{r}{a}\right) - (1+\nu) K^{(9)}\left(\frac{r}{a}\right) K^{(2)Q}\left(\frac{b}{a}\right) \right] + \frac{a^2 \alpha t(1+\nu)}{2h} K^{(9)}\left(\frac{r}{a}\right) + \frac{a^2}{2D} M_1 K^{(6)}\left(\frac{r}{a}\right) + \frac{a^2}{2D} M_2 K^{(8)}\left(\frac{r}{a}\right) \tag{69}$$

$$M_{r\,tot} = \frac{1+\nu}{4\pi} \sum_{j=1}^{n} P_j \left[K^{(2)j}\left(\frac{r}{a}\right) - K^{(10)}\left(\frac{r}{a}\right) K^{(2)j}\left(\frac{b}{a}\right) \right] + (1+\nu)\frac{b}{2} Q \left[K^{(2)Q}\left(\frac{r}{a}\right) - K^{(10)}\left(\frac{r}{a}\right) K^{(2)Q}\left(\frac{b}{a}\right) \right] - \frac{D\alpha t(1+\nu)}{h} \left[1 - K^{(10)}\left(\frac{r}{a}\right) \right] + M_1 K^{(7)}\left(\frac{r}{a}\right) + M_2 K^{(10)}\left(\frac{r}{a}\right) \tag{70}$$

$$M_{t\,tot} = \frac{1+\nu}{4\pi}\sum_{j=1}^{n} P_j\left[K^{(3)j}\left(\frac{r}{a}\right) - K^{(11)}\left(\frac{r}{a}\right)K^{(3)j}\left(\frac{b}{a}\right)\right] + (1+\nu)\frac{b}{2}Q\left[K^{(3)Q}\left(\frac{r}{a}\right) - K^{(11)}\left(\frac{r}{a}\right)K^{(3)Q}\left(\frac{b}{a}\right)\right] - \frac{Da\alpha t(1+\nu)}{h}\left[1 - K^{(11)}\left(\frac{r}{a}\right)\right] + M_1K^{(3)}\left(\frac{r}{a}\right) + M_2K^{(11)}\left(\frac{r}{a}\right) \quad (71)$$

where the summations are over the loads in the unreduced system only.

The magnitude of the radial or tangential bending stresses in the outer fibers can be evaluated from [1],

$$\sigma_r = \frac{6}{h^2}|M_r| \quad (72)$$

$$\sigma_t = \frac{6}{h^2}|M_t| \quad (73)$$

The total unit shear force at any section can be computed by summing all the forces between $r = b$ and the particular radius in question; i.e.

$$2\pi r Q_{tot} = 2\pi\int_b^r q(\xi)\xi d\xi \quad (74)$$

Superposing the elementary problems and bearing in mind the previous definition for total force due to a line load [used with equation (17)], there results

$$Q_{tot} = \frac{b}{r}Q + \sum_{j=2}^{n}\frac{P_j(\xi_j)}{2\pi r} + \frac{P_1}{2\pi r}\int_{b/a}^{r/a}\left(\frac{\xi}{a}\right)\frac{q\left(\frac{\xi}{a}\right)}{q_1}d\left(\frac{\xi}{a}\right) \quad (75)$$

The shear stress on a circumferential line due to this load is maximum at the neutral axis, and can be taken to be [1],

$$\tau = \frac{3}{2}\frac{|Q_{tot}|}{h} \quad (76)$$

It is advisable to make a final check on the slope and deflection boundary conditions. This can be done through the following equations:

Deflection B.C.

$$-(2\pi b)Q(b/a) = w_iK_i = K_i^*w_i^* + \beta P_{tot} \quad (77)$$

$$P_o + P_i = P_{tot} = w_iK_i + w_oK_o \quad (78)$$

Rotation B.C.

$$k_o\left(\frac{dw^*}{dr}\right)_{o\,tot} = (1 - a_{22})M_1 = M_r(1) \quad (79)$$

$$-k_i\left(\frac{dw^*}{dr}\right)_{i\,tot} = b_2 - a_{21}Q - (a_{22} - 1)M_2 = M_r\left(\frac{b}{a}\right) \quad (80)$$

APPLICATIONS

A Summary of Solution Techniques

1 Reduced System

Reduce the general problem of Fig. 6 to the system shown in Fig. 7, using the following equations

$$\beta = \frac{1}{1 - \frac{K_o}{K_i}} \quad (81)$$

$$K_i^* = K_o\frac{1}{1 + \frac{K_o}{K_i}} \quad (82)$$

The final deflection in the actual system is given by

$$w = w^* + \beta\frac{P_{tot}}{K_i} - \beta w_i^* \quad (83)$$

where w^* is the reduced system deflection.

The moments and stresses are the same for both the reduced system and general problem.

2 General Solution Technique for Reduced System

(1) Evaluate and plot $K^{(3)1}$ as a function of position for the distributed load using the graphs for $K_j^{(3)}$ and a graphical or numerical method for performing the integration indicated in equations (14), (15), (16), and (20).

(2) Record all necessary $K^{(3)}$, which are taken from the graphs presented in this section. At this point it is only necessary to take the particular values of $K^{(3)}$ which are necessary to evaluate the a_{ij} and b_i indicated in step 3.

(3) Evaluate all a_{ij} and b_i listed in equation (64).

(4) Evaluate Q, M_1, M_2 from equations (65) to (67) inclusive.

(5) Check this work with equation (63).

(6) Evaluate and plot the deflections and moments using equations (69) to (71) inclusive. These will require the values of $K^{(3)}$ taken from the graphs for different values of r/a.

(7) Evaluate flexural stress at any desired position from equations (72) and (73).

(8) Evaluate the total shear force from equation (75) and the associated shear stress from equation (76).

(9) Check slope and deflection boundary conditions using equations (77) to (80) inclusive.

3 Illustrative Example

The following simple example will serve to illustrate the entire procedure required for the most complex problems.

Given:

Plate, restrained and loaded as in Fig. 9.

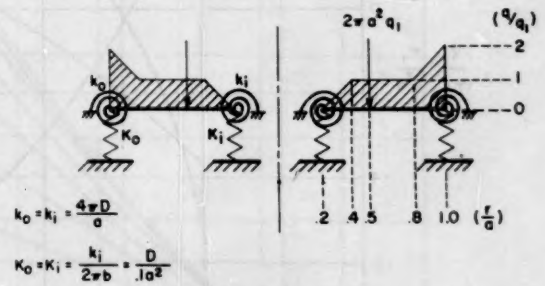


Fig. 9

Find:

Complete moment and deformation solution.

Solution:

1 *Reduced system*

From equations (81) and (82)

$$\beta = \frac{1}{2}, \quad K_i^* = \frac{K_o}{2}$$

then since the total load on the plate is

$$P_{tot} = 1.5466(2\pi a^2 q_1)$$

the reduced system requires a total upward load over the elastic support at $r/a = 0.2$ of $0.7733(2\pi a^2 q_1)$ and is represented in Fig. 10.

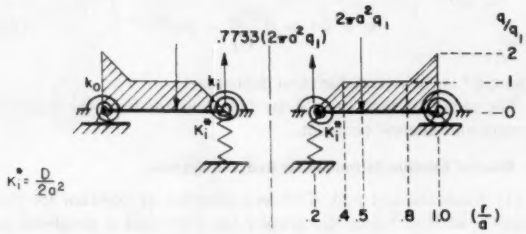


Fig. 10

2 *Solution for reduced system*

Using the graphs and plotting $\left(\frac{q}{q_1}\right)\left(\frac{\xi}{a}\right) K^{(i)}\left(\frac{r}{a}, \frac{\xi}{a}\right)$ against ξ/a, $(i = 1, 2, 3, 4)$ for suitably spaced parameters of r/a and using a planimeter to evaluate each area, there results

r/a	$K^{(1)}$	$K^{(2)}$	$K^{(3)}$	$K^{(4)}$
0.2	0.0411	0.0732	0.0732	−0.038
0.4	0.0330	0.0636	0.0694	
0.6	0.0198	0.0208	0.0487	
0.8	0.0065	−0.0540	0.0106	
1.0	0	−0.1620	−0.0498	

The following additional values of $K^{(i)}\left(\frac{r}{a}\right)$ are obtained directly from the graphs.

$K^{(1),5}(0.2) = 0.189$ $K^{(2),5}(1) = -0.58$ $K^{(4),5}(0.2) = -0.17$
$K^{(1),2}(0.2) = 0.37$ $K^{(2),5}(0.2) = 0.325$ $K^{(4),2}(0.2) = -0.6$
$K^{(6)}(0.2) = 0.625$ $K^{(3),1}(1) = -0.73$ $K^{(7)}(0.2) = 1.505$
$K^{(9)}(0.2) = -0.12$ $K^{(3),2}(0.2) = 1.13$ $K^{(13)}(0.2) = 0.254$
$K^{(10)}(1) = 0.11$
$K^{(12)}(0.2) = -0.723$

In these the second superscript in $K^{(i)}$ indicates the position of load (ξ/a). Using these values in equation (64) the following a_{ij} and b_i matrices are evaluated.

$$a_{ij} = \begin{pmatrix} 2.783\,\frac{ba^2}{2D}\,K_i^* & 0.625\,\frac{a^2}{2D}\,K_i^* & -0.12\,\frac{a^2}{2D}\,K_i^* \\ -0.559b & 10.08 & 0.11 \\ -1.2\pi b & 1.505 & 4.2 \end{pmatrix}$$

$$b_i = \begin{pmatrix} -2.087\,\frac{a^4 q}{2D}\,K_i^* \\ 0.511a^2 q \\ 1.73a^2 q \end{pmatrix}$$

In obtaining b_i, care must be taken to use $P_1 = 2\pi a^2 q_1$, which follows by definition in conjunction with equation (14) to (16) inclusive.

Using these values in equations (65) to (67) inclusive, Q, M_1, and M_2 are evaluated to be,

$$Q = -4.83\,\frac{a^2 q}{2\pi b}$$

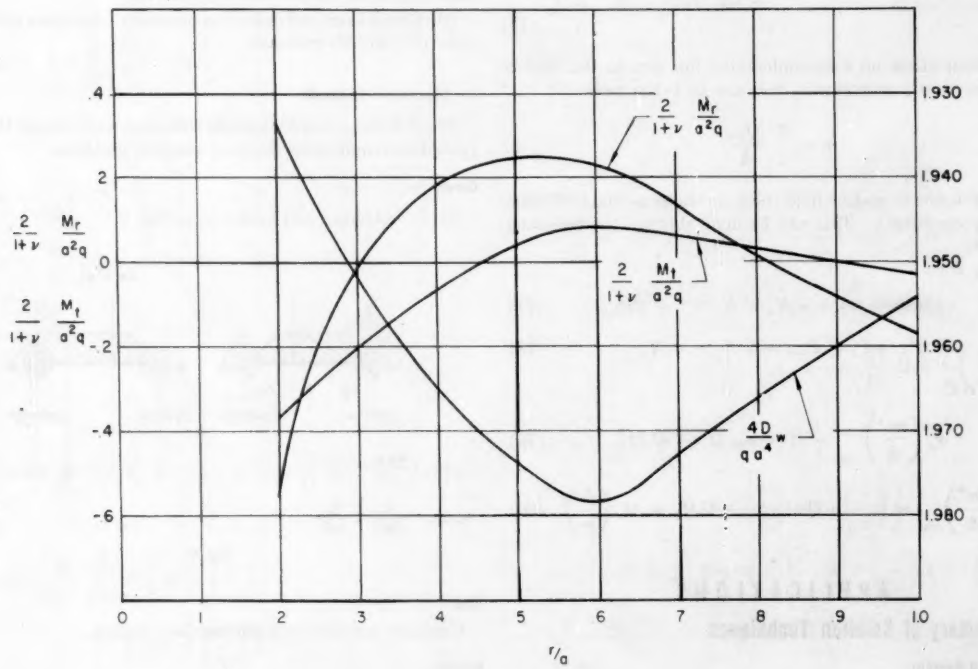


Fig. 11

$$M_1 = 0.0125a^2q$$

$$M_2 = -0.384a^2q$$

As a practical procedure equation (63) should be checked at this stage of a problem before proceeding further.

Substituting these values of Q, M_1, and M_2 in equations (69) to (71) inclusive and reducing them there results,

$$\frac{4D}{qa^4}w^* = K^{(1)1}\left(\frac{r}{a}\right) + K^{(1).5}\left(\frac{r}{a}\right) - 0.769K^{(1).2}\left(\frac{r}{a}\right) + 0.025K^{(6)}\left(\frac{r}{a}\right) - 0.154K^{(9)}\left(\frac{r}{a}\right)$$

$$\frac{2}{1+\nu}\frac{M_r}{a^2q} = K^{(2)1}\left(\frac{r}{a}\right) + K^{(2).5}\left(\frac{r}{a}\right) - 0.769K^{(2).2}\left(\frac{r}{a}\right) + 0.0193K^{(7)}\left(\frac{r}{a}\right) - 0.12K^{(10)}\left(\frac{r}{a}\right)$$

$$\frac{2}{1+\nu}\frac{M_t}{a^2q} = K^{(3)1}\left(\frac{r}{a}\right) + K^{(3).5}\left(\frac{r}{a}\right) - 0.769K^{(3).2}\left(\frac{r}{a}\right) + 0.0193K^{(8)}\left(\frac{r}{a}\right) - 0.12K^{(11)}\left(\frac{r}{a}\right)$$

From equation (83) the actual deflection is,

$$\frac{4D}{qa^4}w = 1.9546 + \frac{4D}{qa^4}w^*$$

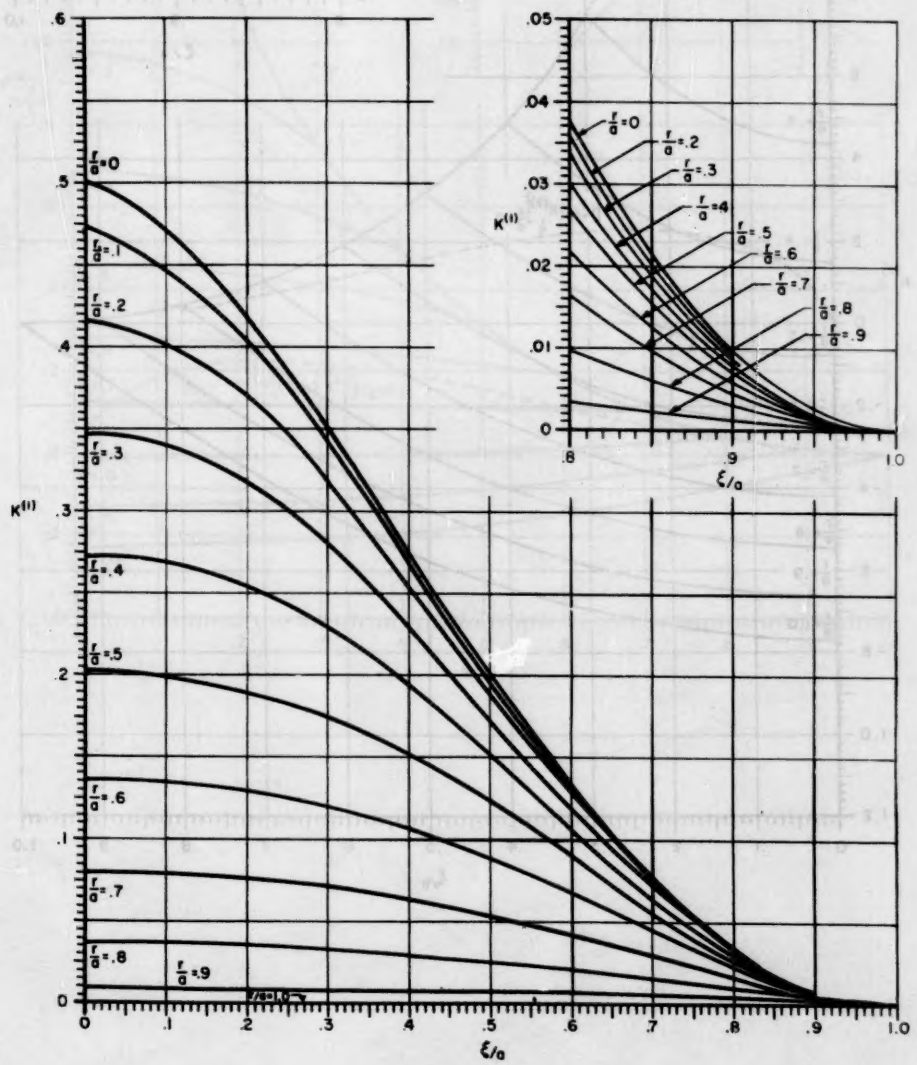
These solutions are shown in Fig. 11.

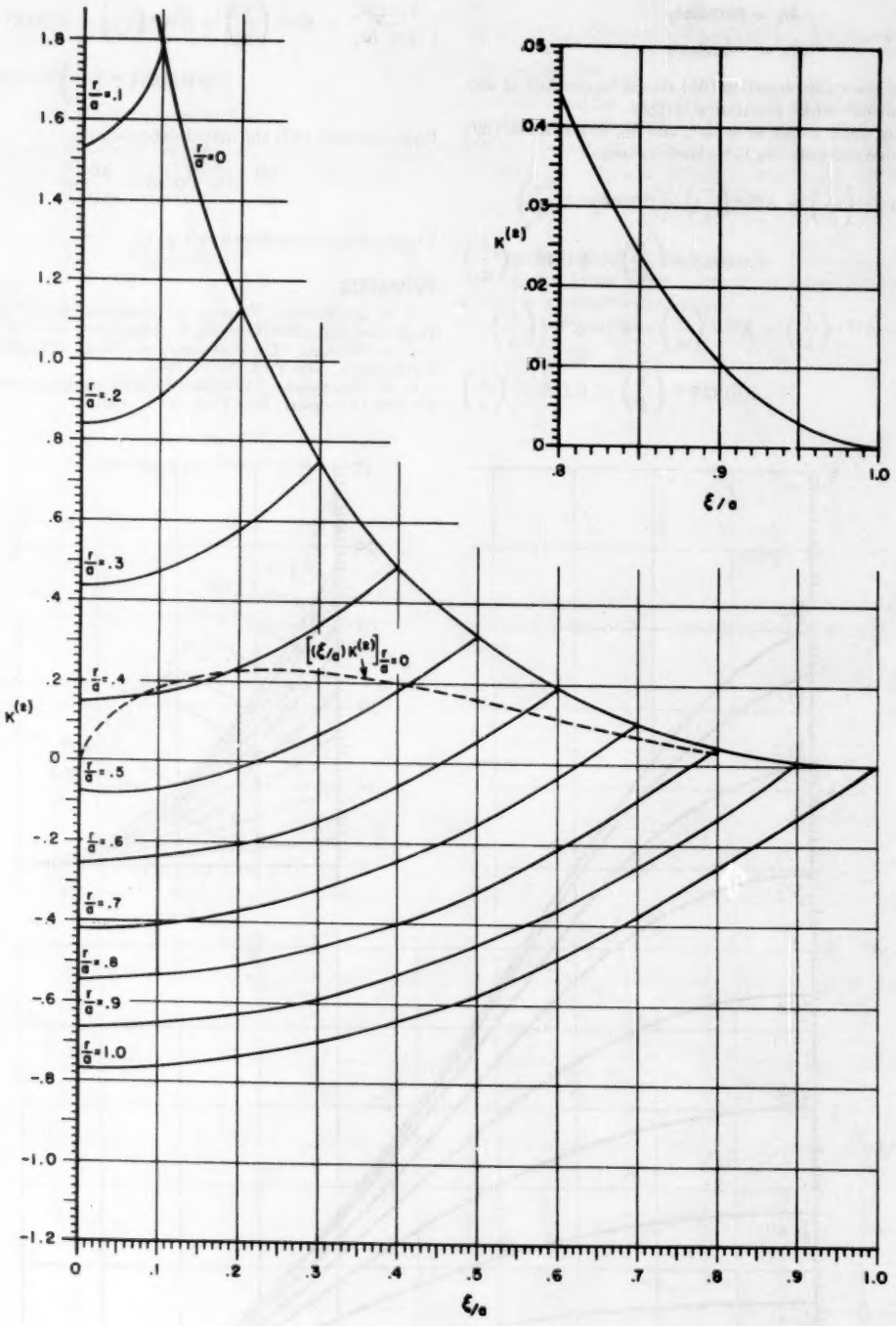
References

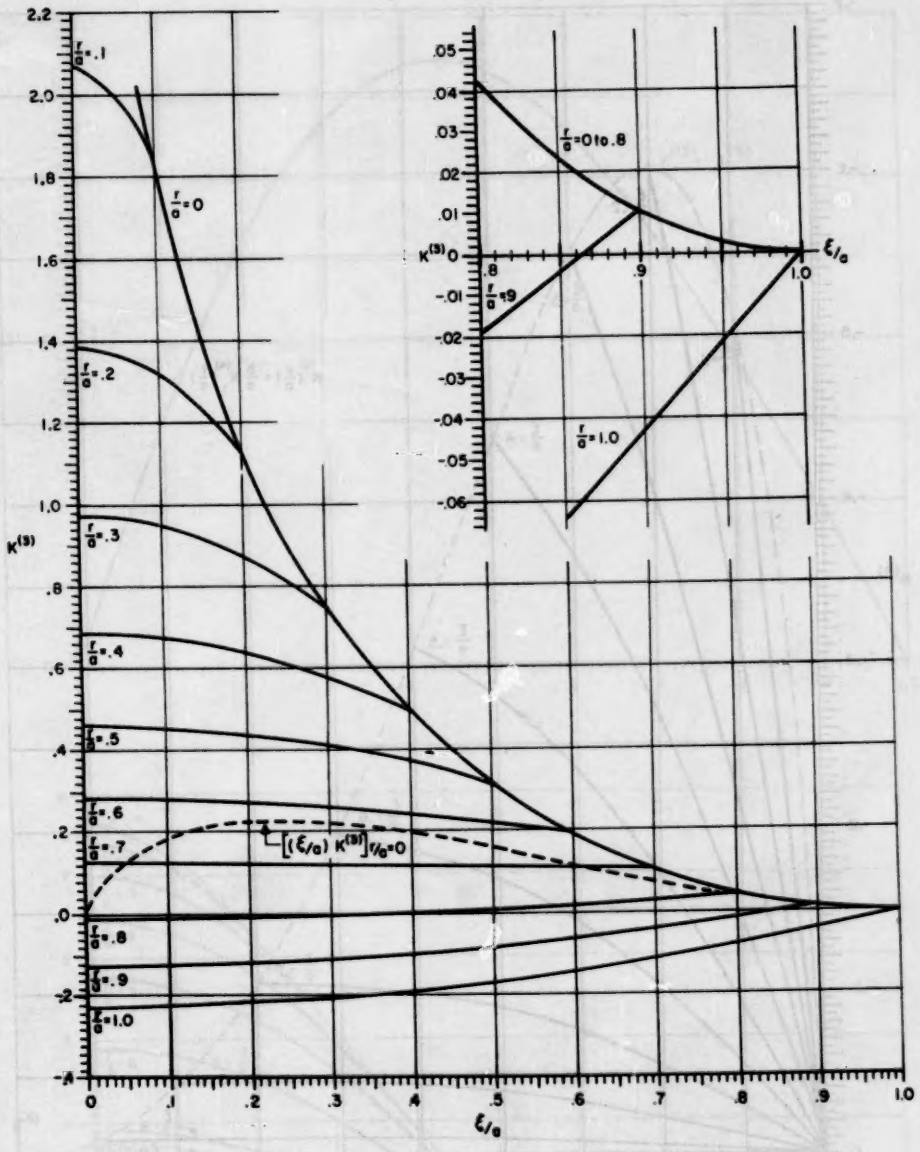
1 S. Timoshenko, "Theory of Plates and Shells," McGraw-Hill Book Company, New York, N. Y., 1940.

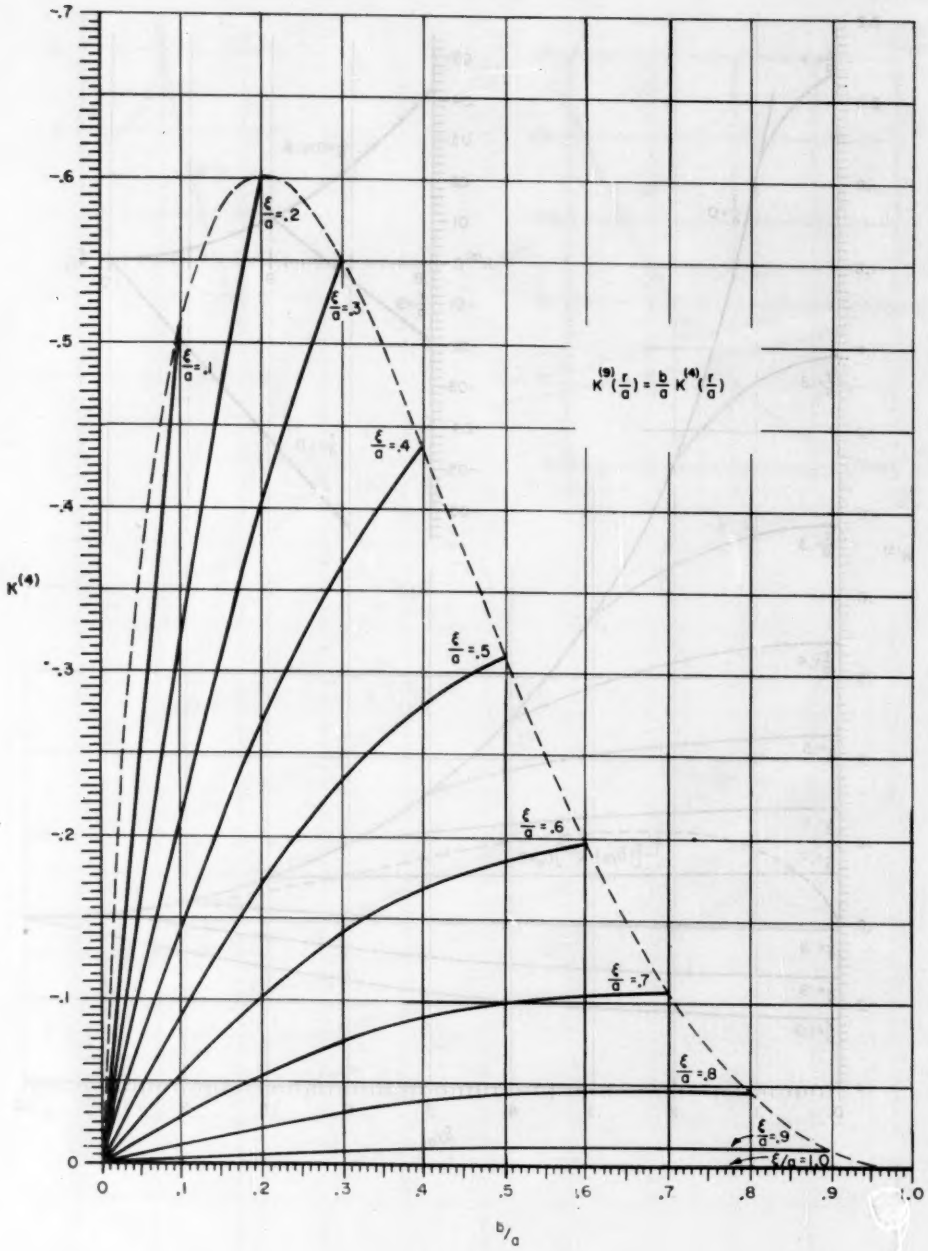
2 A. E. Love, "The Mathematical Theory of Elasticity," Dover Publications, New York, N. Y., 1944.

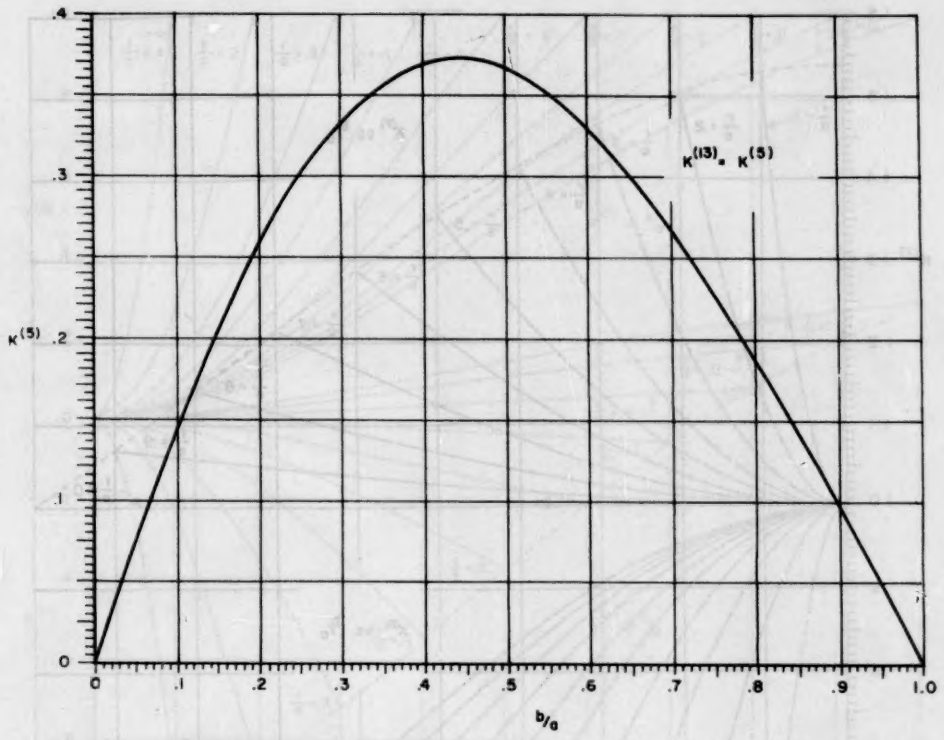
3 S. Timoshenko, "Vibration Problems in Engineering," D. Van Nostrand Company, New York, N. Y., 1937.

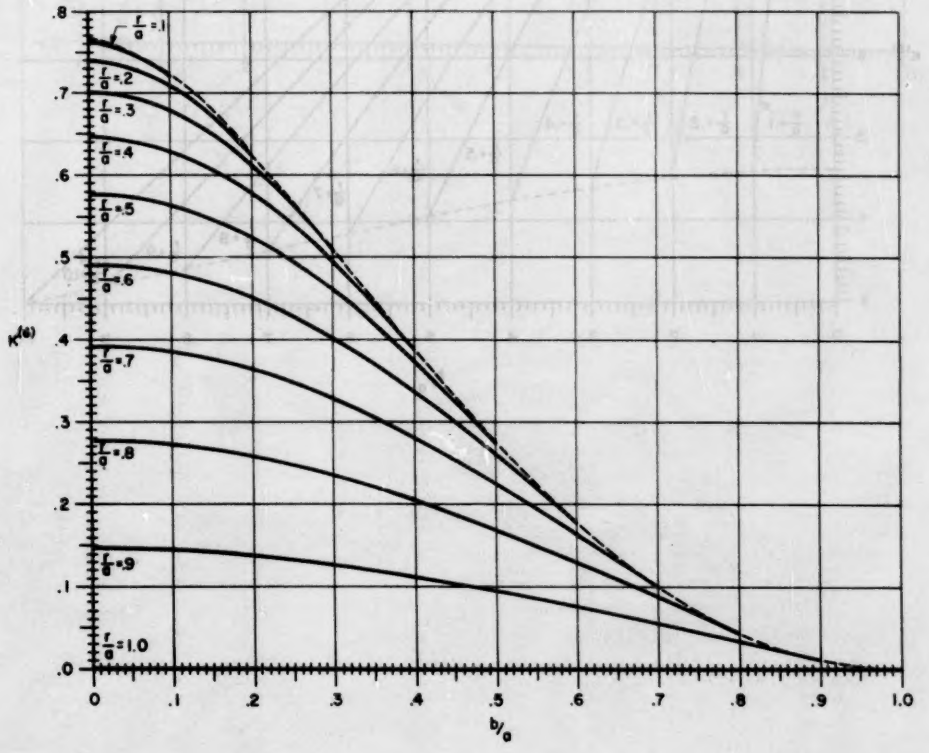


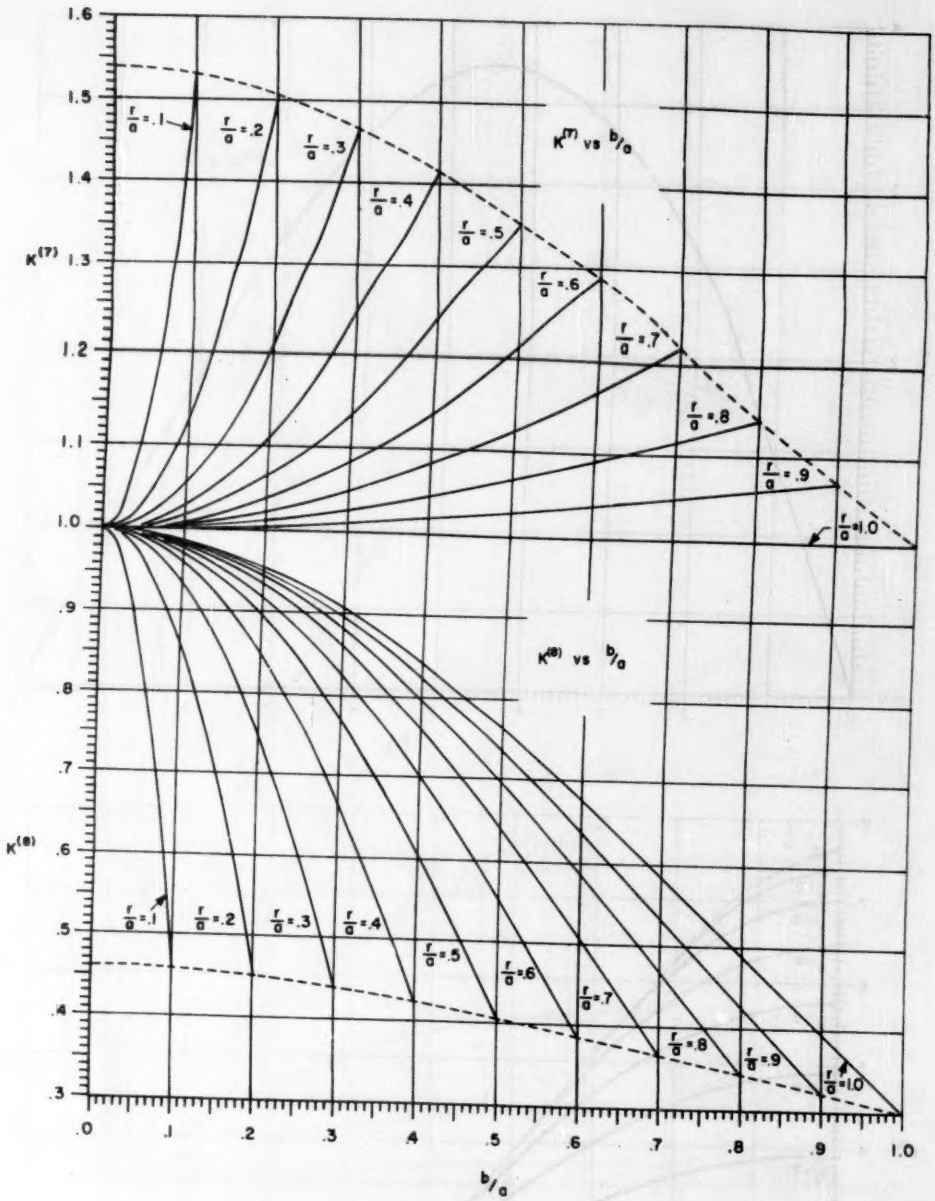

1.8
1.6
1.4
1.2
1.0
.8
.6
.4
.2
0
-.2
-.4
-.6
-.8
-1.0
-1.2
K(ε)
r/a=.1
r/a=0
r/a=.2
r/a=.3
r/a=.4
r/a=.5
r/a=.6
r/a=.7
r/a=.8
r/a=.9
r/a=1.0
[(ξ/a)K(ε)]r/a=0
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1.0
ξ/a
.05
.04
.03
.02
.01
0
K(ε)
.8
.9
1.0
ξ/a

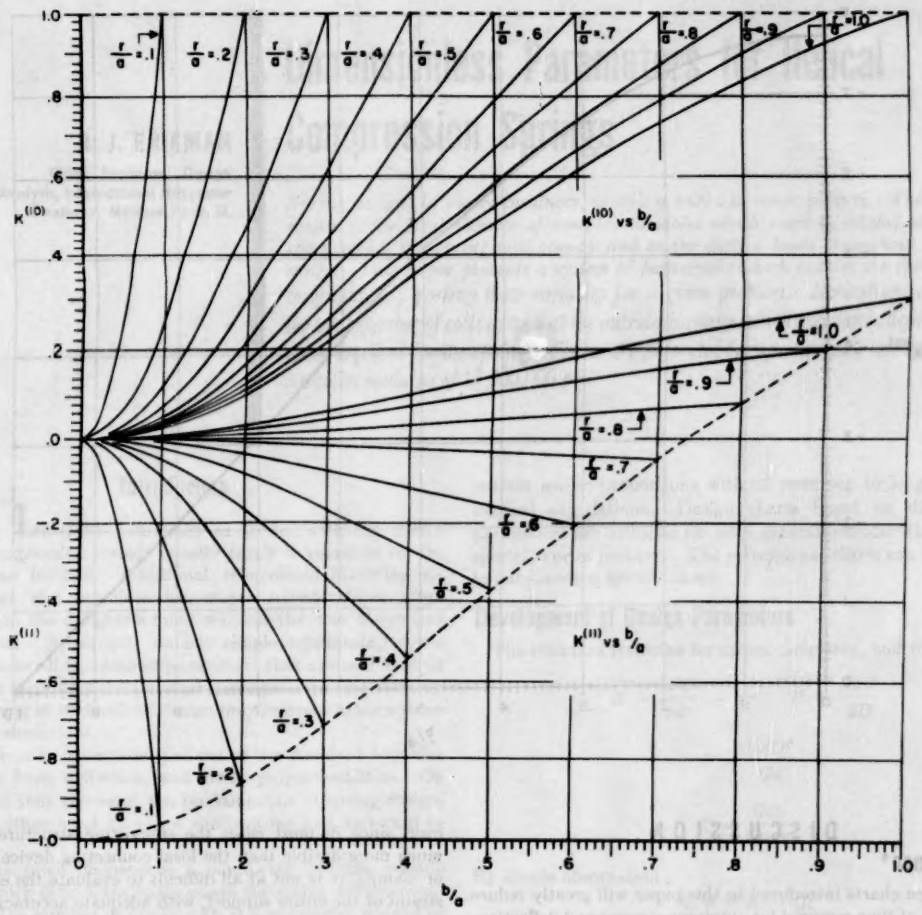

$K^{(3)}$
ξ/a
$\frac{r}{a}$=.1
$\frac{r}{a}$=0
$\frac{r}{a}$=.2
$\frac{r}{a}$=.3
$\frac{r}{a}$=.4
$\frac{r}{a}$=.5
$\frac{r}{a}$=.6
$\frac{r}{a}$=.7
$\frac{r}{a}$=.8
$\frac{r}{a}$=.9
$\frac{r}{a}$=1.0
$[(\xi/a)\,K^{(3)}]_{r/a=0}$
$\frac{r}{a}$=0 to .8
$\frac{r}{a}$=.9
$\frac{r}{a}$=1.0

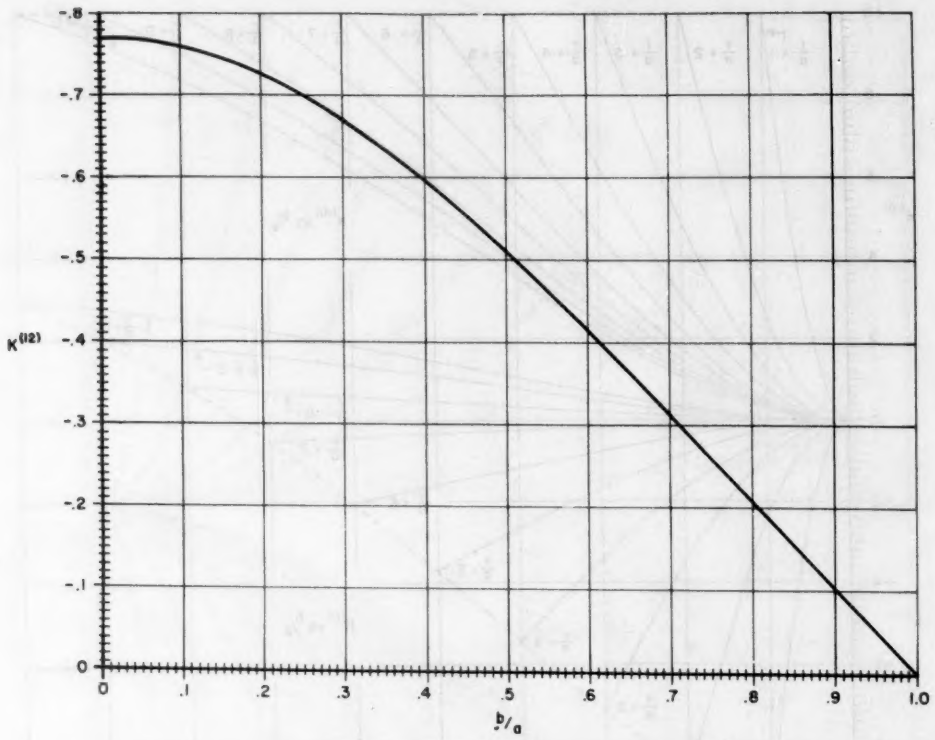

$K^{(9)}(\frac{r}{a}) = \frac{b}{a} K^{(4)}(\frac{r}{a})$
$K^{(4)}$
b/a
$\frac{\xi}{a}=.1$
$\frac{\xi}{a}=.2$
$\frac{\xi}{a}=.3$
$\frac{\xi}{a}=.4$
$\frac{\xi}{a}=.5$
$\frac{\xi}{a}=.6$
$\frac{\xi}{a}=.7$
$\frac{\xi}{a}=.8$
$\frac{\xi}{a}=.9$
$\xi/a=1.0$
-.7
-.6
-.5
-.4
-.3
-.2
-.1
0
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1.0


$K^{(13)} = K^{(5)}$
$K^{(5)}$
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1.0
b/a


$\frac{t}{a}=.1$
$\frac{t}{a}=.2$
$\frac{t}{a}=.3$
$\frac{t}{a}=.4$
$\frac{t}{a}=.5$
$\frac{t}{a}=.6$
$\frac{t}{a}=.7$
$\frac{t}{a}=.8$
$\frac{t}{a}=.9$
$\frac{t}{a}=1.0$
$K^{(6)}$
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1.0
b/a


$K^{(T)}$ vs b/a
$K^{(B)}$ vs b/a
$K^{(T)}$
$K^{(B)}$
b/a
$\frac{r}{a}=.1$
$\frac{r}{a}=.2$
$\frac{r}{a}=.3$
$\frac{r}{a}=.4$
$\frac{r}{a}=.5$
$\frac{r}{a}=.6$
$\frac{r}{a}=.7$
$\frac{r}{a}=.8$
$\frac{r}{a}=.9$
$\frac{r}{a}=1.0$
1.6
1.5
1.4
1.3
1.2
1.1
1.0
.9
.8
.7
.6
.5
.4
.3
.0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1.0


1.0
.8
.6
.4
.2
.0
-.2
-.4
-.6
-.8
-1.0
K(10)
K(11)
0
.1
.2
.3
.4
.5
.6
.7
.8
.9
1.0
b/a
r/a = .1
r/a = .2
r/a = .3
r/a = .4
r/a = .5
r/a = .6
r/a = .7
r/a = .8
r/a = .9
r/a = 1.0
K(10) vs b/a
K(11) vs b/a



DISCUSSION

R. J. Erisman[4]

The design charts introduced in this paper will greatly reduce the amount of time required for obtaining stresses and deflections in circular plates. These data are particularly valuable since they are presented in forms which can be readily adapted to conditions having varying degrees of loading and constraint. A problem which still confronts the designer is the determination of the amount of constraint that is present by the usual fabrication methods of welding, clamping, or bolting. Also, the designer, in many cases, must deviate from rigorous mathematical solutions to account for a change in plate thickness. Perhaps the authors can contribute simplifications to this phase of the overall problem also.

Authors' Closure

The authors wish to thank Mr. Erisman for his interesting comments. The problem of the *precise* evaluation of constraint associated with the indicated connecting methods is quite involved and normally not suitable for design calculations. However, since in most cases the supporting structure is in itself much more flexible than the local connecting device (bolt, weld, or clamp), it is not at all difficult to evaluate the effective constraint of the entire support, with adequate accuracy, by known structural techniques. In the event that this is not convenient, one can resort to experimental procedures which establish the effective rigidity of a supporting structure. In fact, one can establish a good combined experimental-analytical set of graphs that would provide coefficients for the support restraint to supplement the graphs of this paper.

The problem of a stepped plate is easily taken care of within the framework of the techniques presented in the paper. In that case one need only treat each stepped element as a separate plate subject to force-equilibrium and compatibility conditions at the stepped section. On the other hand, the case of the continuously varying plate has not been treated in this paper. This could be taken care of in one of the following ways. First, one could approximate it as a stepped plate with a suitable number of steps (in much the same way as variable diameter shafting is often analyzed), and then use the technique discussed above. Second, one could expand the curves presented in the paper to include general families of variable thickness plates. This second task has not been undertaken and would seem to serve a useful purpose in rounding out the presented work.

[4] Chief Engineer, Design Analysis Section, Construction Equipment Engineering Department, International Harvester Company, Melrose Park, Ill.

Dimensionless Parameters for Helical Compression Springs

R. J. ERISMAN
Chief Engineer, Design Analysis, International Harvester Company, Melrose Park, Ill.

Spring design, in many instances, is still a trial-and-error process. The underlying reason is the large number of complex variables which must be related to the allotted space which the spring will occupy and to the desired loads throughout the working range. This paper presents a system of parameters which enables the spring designer to graphically portray these variables for a given problem. Derivations are based on the total number of coils as well as the outside diameter which accurately defines the entire space required for the spring. Charts are presented for springs with two dead coils and a torsion modulus of 11,500,000 psi.

Introduction

THE MANY VARIABLES associated with the design of helical compression springs usually result in solutions by the trial-and-error method. Additional refinements involving restricted space and optimum load-stress considerations often leave doubt in the designer's mind whether the best design has been achieved. Apparently no one simple relationship exists which expresses all the desired parameters that are encountered in the design of this otherwise simple machine element. As a result, any spring in its finalized design may represent many time-consuming calculations.

These difficulties stem from the use of the standard formulas which relate load, deflection, and stress proportionalities. On the one hand they represent the fundamentals of spring design, but on the other hand they are cumbersome and awkward to handle.

First: The pitch diameter is a parameter, not the outside diameter, which defines the space where the spring will operate.

Second: The number of active coils is a parameter instead of the total coils or its related solid height.

Third: Numbers must be raised to the second, third, and fourth powers to evaluate the formulas.

The purpose of this paper is to present a means of determining correct spring proportions without resorting to lengthy mathematical calculations. Design charts based on dimensionless parameters are included for both general considerations and one special type of problem. The principles outlined can be extended to handle other specific cases.

Contributed by the Machine Design Division and presented at the Annual Meeting, Atlantic City, N. J., November 29–December 4, 1959, of THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS.

NOTE: Statements and opinions advanced in papers are to be understood as individual expressions of their authors and not those of the Society. Manuscript received at ASME Headquarters, May 27, 1959. Paper No. 59—A-30.

Development of Design Parameters

The standard formulas for stress, deflection, and rate are:

$$S = \frac{8PD}{\pi d^3} \quad \text{or} \quad P = \frac{S\pi d^3}{8D} \tag{1}$$

$$F = \frac{S\pi D^2}{Gd} \tag{2}$$

$$R = \frac{Gd^4}{8D^3N} \tag{3}$$

By simple observation

$$N = (CK - Q) \tag{4}$$

Introducing this term into equation (3) together with the other indicated substitutions for the diameters d and D results in:

$$\frac{R}{GD_o} = \frac{1}{8K(K-1)^3(CK-Q)}$$

$$\text{or} \quad \frac{R}{D_o} = \frac{G}{8K(K-1)^3(CK-Q)} \tag{5}$$

When G and Q are assigned values, a family of curves can be drawn to represent this relationship. Fig. 1 shows such a graph

Nomenclature

C = parameter for length (see Dimensionless Parameters)
d = wire diameter, in.
D = mean diameter of spring, in.
D_o = outside diameter of spring, in.
f = deflection from test load to final load, in.
f_1 = deflection from free length to test load, in.
F = deflection per coil, in.
G = torsional modulus of elasticity, psi
H = solid height of spring, in.
K = parameter for diameter (see Dimensionless Parameters)
K_w = Wahl stress correction factor, $K_w = \left[\frac{K-1.25}{K-2} + \frac{0.615}{K-1}\right]$
L = free length, in.
N = number of active coils
P = test load (initial spring load), lb
Q = number of dead coils
R = rate of spring, lb/in.
S = torsional stress, psi
S' = parameter for stress, $S' = \frac{SD_o^2}{P}$ (see Dimensionless Parameters)
W = final spring load, lb

Dimensionless Parameters

$C = H/D_o$
$K = D_o/d$

$$\frac{R}{GD_o} = \frac{1}{8K(CK-Q)(K-1)^3}$$

$$\frac{SD_o^2}{P} = \frac{8}{\pi}(K-1)K^2$$

$$\frac{SD_o}{f_1G} = \frac{K}{\pi(CK-Q)(K-1)^2}$$

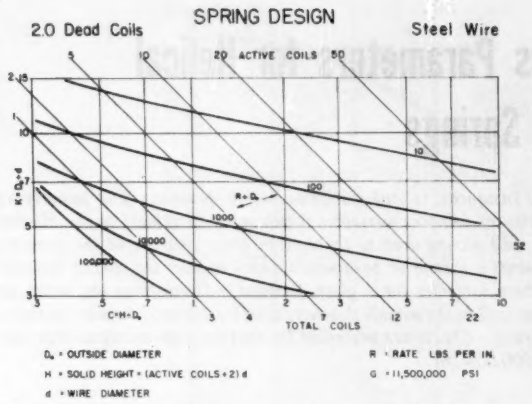

SPRING DESIGN
2.0 Dead Coils
Steel Wire
20 ACTIVE COILS 50
TOTAL COILS
D₀ = OUTSIDE DIAMETER
H = SOLID HEIGHT = (ACTIVE COILS + 2) d
d = WIRE DIAMETER
R = RATE LBS PER IN
G = 11,500,000 PSI

Fig. 1 Design chart with parameters for diameter, height, rate, and number of coils

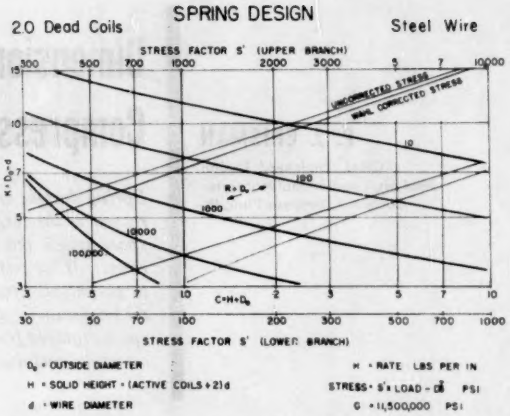

SPRING DESIGN
2.0 Dead Coils
Steel Wire
STRESS FACTOR S' (UPPER BRANCH)
UNCORRECTED STRESS
WAHL CORRECTED STRESS
STRESS FACTOR S' (LOWER BRANCH)
D₀ = OUTSIDE DIAMETER
H = SOLID HEIGHT = (ACTIVE COILS + 2) d
d = WIRE DIAMETER
R = RATE LBS PER IN
G = 11,500,000 PSI

Fig. 2 Design chart with parameters for diameter, height, rate, and stress

for $G = 11{,}500{,}000$ psi and $Q = 2$. It will be noticed that the number of coils (active or total) can also be represented on the same co-ordinate system. If sufficient R/D_o lines are drawn, the chart will instantly show what size relationships will provide a given rate.

Stress parameters can be developed from equation (1) in a manner similar to that used by Hinkle and Morse [1].[1]

$$\frac{SD_o^2}{P} = \frac{8}{\pi}(K-1)K^2 \text{ (uncorrected)} \tag{6}$$

$$\frac{SD_o^2}{P} = \frac{8}{\pi}(K-1)K^2\left[\frac{K-1.25}{K-2} + \frac{0.615}{K-1}\right] \text{ (Wahl corrected)} \tag{7}$$

Also since

$$P = Rf_1 \tag{8}$$

$$\frac{SD_o}{f_1 G} = \frac{K}{\pi(K-1)^2(CK-Q)} \text{ (uncorrected)} \tag{9}$$

$$\frac{SD_o}{f_1 G} = \frac{K}{\pi(K-1)^2(CK-Q)}\left[\frac{K-1.25}{K-2} + \frac{0.615}{K-1}\right] \text{ (Wahl corrected)} \tag{10}$$

A useful design chart can be made by superimposing stress data with the constant rate lines of Fig. 1. This has been done in Fig. 2, again using $G = 11{,}500{,}000$ psi and $Q = 2$. The stress relationship given in equations (6) and (7) was chosen because, in most cases, the load P will be known before the corresponding deflection f_1.

Example: 1 (Using Fig. 2)

Given data: $W = 45$ lb @ 1.75 in.
$S = 85{,}000$ psi (Wahl corrected)
$D_o = 1.25$ $\quad H = 1.5$

Required: d, L

Solution: Calculate stress parameter

$$S' = \frac{S \times D_o^2}{W} = \frac{85{,}000 \times \overline{1.25}^2}{45} = 2950$$

Calculate height parameter

$C = H/D_o = 1.5/1.25 = 1.2$.

At intersection of $S' = 2950$ and upper branch of Wahl corrected stress line read $K = 10.3$.

Hence, $d = D_o/K = 1.25/10.3 = 0.121$ in.

Proceed horizontally to the left until intersecting $C = 1.2$. Read $R/D_o = 16$.

Hence, $R = 16 \times 1.25 = 20$ lb/in.

Deflection $= W/R = 45/20 = 2.25$ in.

Free length $= 1.75 + 2.25 = 4.0$ in.

Principles Applied to a Specific Problem

Consider the problem of designing a spring for a given test load P with additional deflection f for minimum stress at final load W. Or, stated as a corollary, what maximum load P can be specified for a given allowable stress at final load W? Both solid height and outside diameter are specified.

Applying conventional methods to this problem, the spring designer would obtain stresses for several different wire sizes to obtain a stress diagram as in Fig. 3. If the minimum stress is conservative, a greater P load would be chosen until the minimum stress point is at the maximum limit for the material chosen.

Jennings [2] developed equations and charts for the solution of this problem. His analysis, however, did not include the effect of dead coils. Also, the pitch diameter was used for the size parameter and the final working length was at solid height.

Using the principles outlined thus far, these limitations can be overcome. From equations (5) and (6), the uncorrected stress at final load is

$$S = \frac{8}{\pi}(K-1)\frac{K^2}{D_o^3}\left[P + \frac{GfD_o}{8K(K-1)^2(CK-Q)}\right] \tag{11}$$

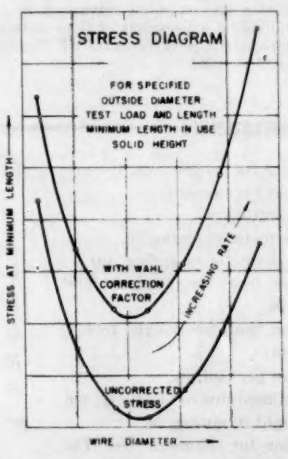


Fig. 3 Diagram showing minimum stress condition for data as specified

[1] Numbers in brackets designate References at end of paper.

When equation (11) is differentiated with respect to K, the condition for minimum stress is found to be

$$\frac{8P}{GfD_o} = \frac{2CK^3 - Q(K+1)}{K(3K-2)(CK-Q)^2(K-1)^2} \tag{12}$$

Again, this relationship can be plotted as a family of curves to form a design chart, see Fig. 4. This chart enables a spring designer to immediately select the wire size which produces the minimum stress from known size and loading considerations.

Using equations (5) and (7) the stress with the Wahl correction factor at final load is obtained.

$$S = \frac{8}{\pi}(K-1)\frac{K^2}{D_o^2}\left[P + \frac{GfD_o}{8K(CK-Q)(K-1)^2}\right]\left[\frac{K-1.25}{K-2} + \frac{0.615}{K-1}\right] \tag{13}$$

Differentiating this equation with respect to K, the condition for minimum stress is found to be

$$\frac{8P}{GfD_o} = \frac{2CK^5 - (5.90C+Q)K^4 + (2.62C+1.27Q)K^3 + (3.13C+4.30Q)K^2 - 6.46QK + 0.04Q}{K(CK-Q)^2(K-1)^4(3K^2 - 11.27K + 9.83K - 0.08)} \tag{14}$$

The plot of these data is represented in Fig. 5.

These data can be presented in another form, namely, the ratio of W to P which produces the condition of minimum stress. This is developed as follows (for the uncorrected stress condition):

$$P = Rf_1 \tag{8}$$

$$W = R[f_1 + f] \tag{15}$$

$$\frac{W}{P} = 1 + \frac{f}{f_1} \tag{16}$$

combining equation (5), (12), (8), and (16) gives

$$\frac{W}{P} = 1 + \frac{(CK-Q)(3K-2)}{2CK^2 - Q(K+1)} \tag{17}$$

Cross plotting this equation with equation (12) (to eliminate the K term) results in the graph shown in Fig. 6. Here it is a simple matter to determine the required final load for minimum stress from the size limitations imposed, the initial load, and supplementary deflection.

In order to obtain the maximum value of P consistent with an allowable stress at W, one additional design chart is required. It is developed for the case of the uncorrected stress from an extension of equation (9)

$$\frac{SD_o}{G} = \frac{K}{\pi(CK-Q)(K-1)^2}[f_1 + f] \tag{18}$$

from equations (16) and (17)

$$f_1 = f\left[\frac{2CK^2 - Q(K+1)}{(3K-2)(CK-Q)}\right] \tag{19}$$

and finally from a combination of equations (18) and (19)

$$\frac{SD_o}{Gf} = \frac{K[(CK-Q)(5K-2) + Q(K-1)]}{\pi(K-1)^2(CK-Q)^2(3K-2)} \tag{20}$$

Fig. 7 is a plot of these data based on $G = 11{,}500{,}000$ psi and $Q = 2$. It is superimposed on the data shown in Fig. 4. In addition, the R/D_o lines are shown to make a single chart with as many of the design variables as possible.

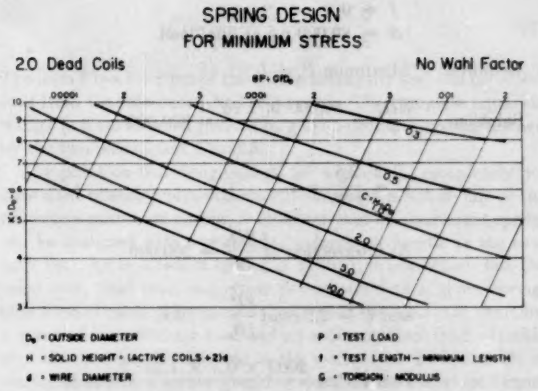

SPRING DESIGN
FOR MINIMUM STRESS
20 Dead Coils
No Wahl Factor
D_o = OUTSIDE DIAMETER
H = SOLID HEIGHT = (ACTIVE COILS + 2)d
d = WIRE DIAMETER
P = TEST LOAD
f = TEST LENGTH − MINIMUM LENGTH
G = TORSION MODULUS

Fig. 4 Design chart for obtaining wire diameter to produce minimum uncorrected stress

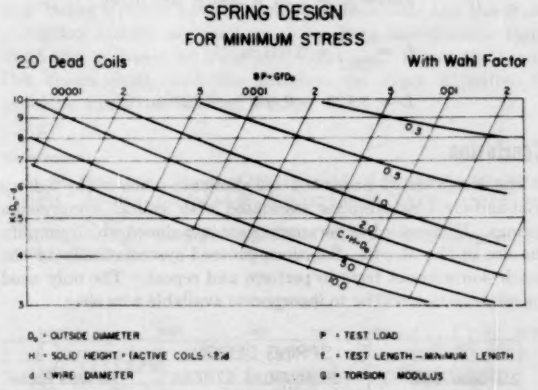

SPRING DESIGN
FOR MINIMUM STRESS
20 Dead Coils
With Wahl Factor
D_o = OUTSIDE DIAMETER
H = SOLID HEIGHT = (ACTIVE COILS + 2)d
d = WIRE DIAMETER
P = TEST LOAD
f = TEST LENGTH − MINIMUM LENGTH
G = TORSION MODULUS

Fig. 5 Design chart for obtaining wire diameter to produce minimum Wahl corrected stress

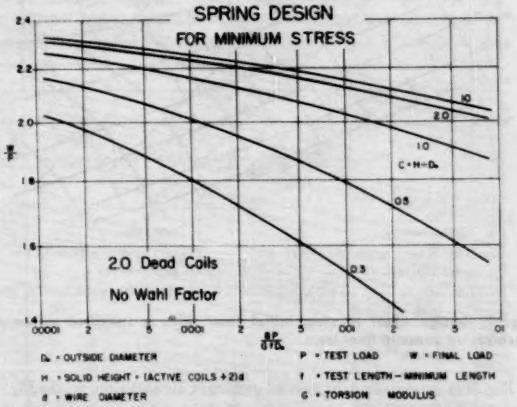

SPRING DESIGN
FOR MINIMUM STRESS
20 Dead Coils
No Wahl Factor
D_o = OUTSIDE DIAMETER
H = SOLID HEIGHT = (ACTIVE COILS + 2)d
d = WIRE DIAMETER
P = TEST LOAD
W = FINAL LOAD
f = TEST LENGTH − MINIMUM LENGTH
G = TORSION MODULUS

Fig. 6 Design chart for determining what final load is required to produce minimum uncorrected stress

Example: 2 (Using Fig. 7)

Given data: $D_o = 1.25$ Length at initial load = 3.125

$H = 2.5$ Length at final load = 2.625

$f = 0.5$
$S = 80{,}000$ psi at final load

Required: Maximum P, d, L

Solution: $\dfrac{SD_o}{f} = \dfrac{80{,}000 \times 1.25}{0.5} = 200{,}000$

$C = \dfrac{2.5}{1.25} = 2$

At intersection of $\dfrac{SD_o}{f} = 200{,}000$

and $C = 2.0$ read $\dfrac{8P}{fD_o} = 2000.$

Hence $P = \dfrac{2000 \times 0.5 \times 1.25}{8} = 156$ lb.

Also read $K = 5.57$.

Hence $d = \dfrac{1.25}{5.57} = 0.224$ in.

Also read $R/D_o = 306$.
Hence $R = 1.25 \times 306 = 382$ lb/in.

$f_1 = \dfrac{156}{382} = 0.408$ in.

$L = 3.125 + 0.408 = 3.533$ in.

Conclusion

The design charts presented will eliminate much of the tedious trial-and-error calculations associated with helical compression springs. Dimensionless parameters are introduced which simplify the use of the charts. Solutions proceed systematically to the conclusion without tests to perform and repeat. The only need for recalculation will be to incorporate available wire sizes.

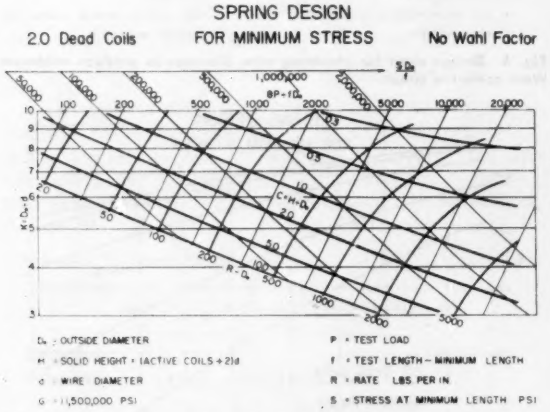


Fig. 7 Design chart relating initial load with a specified minimum uncorrected stress at final load

Fig. 2 is especially adapted to general calculations, while Fig. 7 represents the solution to one particular problem. A spring designer can apply these principles to other problems and obtain similar simplifications.

References

1 R. T. Hinkle and I. E. Morse, Jr., "Design of Helical Springs for Minimum, Weight, Volume, and Length," TRANS. ASME, Series B, JOURNAL OF ENGINEERING FOR INDUSTRY, vol. 81, 1959, p. 37.

2 J. Jennings, "Maximum Capacity of Helical Springs," *Machinery* (London, England), vol. 51, 1937, p. 109.

DISCUSSION

P. O. Bell[2]

The author has developed a valuable method of determining correct spring proportions for the design of helical compression springs. He introduces a system of spring variables which are related to the desired working loads and allotted space to be occupied by the spring, and which includes the outside diameter of the spring instead of the pitch diameter, the total number of coils (or its related height) instead of only the number of active coils. In terms of these variables and the standard formulas for stress, deflection, and rate, a system of dimensionless parameters is introduced. The mathematical relations among these parameters may be represented graphically by the spring designer to serve as "Design charts" for a given problem. Such design charts, some of which are presented in this paper, are great labor saving devices, which will eliminate, for the spring designer, most of the tedius trial-and-error calculations associated with the design of helical compression springs.

M. Gerard Fangemann[3]

The author of this paper has made a very useful contribution to the practical solution of compression spring design problems. The standard formulas involving the mean coil diameter and the number of active coils are recast, using more practical spring dimensions, namely, outside diameter and solid height. Simplification is achieved by the use of several ratios, called parameters for length, diameter, and stress and the plotting of curves using these parameters as ordinates. Several examples are worked out showing that spring dimensions can be obtained from the curves with great rapidity.

The curves presented in this paper are intended for steel compression springs, made of round wire with ground ends and having two dead coils. This means that similar curves must be worked out for various other possibilities, such as materials other than steel, square or rectangular wire, plain ends, and less than two dead coils. Thus it would be necessary for the spring specialist to provide himself with many charts of curves. In addition, it would appear desirable to have more curves for each chart to permit a more accurate determination of the co-ordinates to be read. Due to the fact that compression spring design is often very tedious and time consuming due to the trial-and-error method required by the conventional formulas, the spring designer may find it quite profitable in time and effort to make the charts as described.

C. H. Tsao[4]

The author is to be congratulated for taking another step forward in the field of spring design. His paper offers a simple and direct method of determining the best spring proportions. The design charts are especially useful for all helical spring designers.

---

[2] Senior Staff Physicist, Systems Analysis Laboratory, Hughes Aircraft Company, Culver City, Calif.

[3] General Manager, Spring Division, John Chatilion & Sons, New York, N. Y. Mem. ASME.

[4] Member, Technical Staff, Guided Missile Laboratories, Hughes Aircraft Company, Culver City, Calif.

Author's Closure

The author wishes to thank each of the discussers for the comments submitted. Mr. Fangemann correctly indicates the need of many curves on a particular design chart before it can be considered a practical design tool. Unfortunately, material prepared for publication is usually abbreviated for clarity, since understanding of the method is the primary consideration. The chart we have prepared for our own use, corresponding to Fig. 2, is appended herewith for more accurate computations. To extend the use of the chart for different moduli of elasticity, the parameter R/D_o can be expanded to the equivalent $\frac{R}{D_o}\frac{(11{,}500{,}000)}{G'}$.

This chart contains three additional functions which will be useful to the spring designer. These functions give the geometric relationships for minimum weight, minimum volume, and minimum length. Derivations are based on the substitution of equations (5) and (7) in the design constant given by Hinkle and Morse [1]:

$$B = \frac{f_1 G}{Q(8P\pi S)^{1/2}} \tag{21}$$

Rearranging this equation and the others referred to, results in:

$$\left(8\pi \frac{S}{P}\right)^{1/2} \frac{P}{f_1} = \frac{G}{BQ} \tag{21a}$$

$$\pi \frac{S}{P} = \frac{8(K-1)K^2K_w}{D_o^2} \tag{7}$$

$$R = \frac{P}{f_1} = \frac{G\,D_o}{8K(K-1)^3(CK-Q)} \tag{5}$$

$$C = \frac{BQK_w^{1/2}}{(K-1)^{3/2}K} + \frac{Q}{K} \tag{22}$$

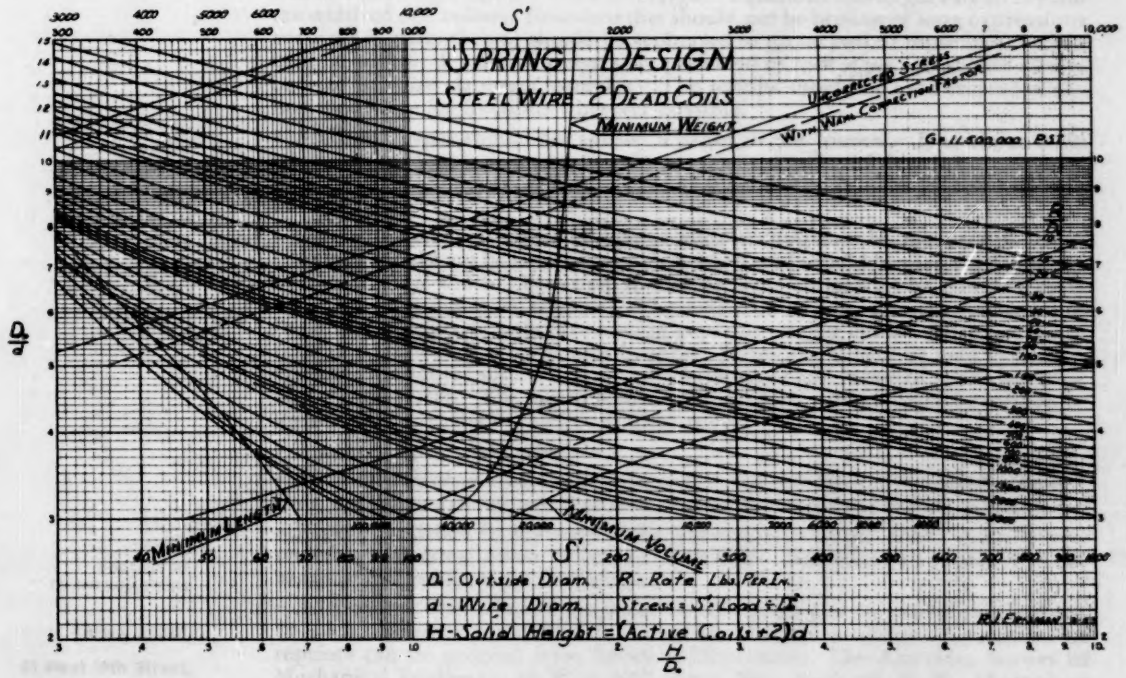
The term B is a function of the spring index only and can be evaluated from the referenced ASME paper. The complete equation results in a curve when plotted on the co-ordinate system defined by the two parameters C and K.

Any point on this design chart, for a given D_o, completely defines a set of spring proportions. If the point is not on one of the respective minimum curves, it indicates that an equivalent spring can be designed with less weight, volume, or length as the case may be. An equivalent spring is defined as one which has the same rate, final load, and stress as compared to a given spring. One word of caution is necessary regarding the use of the chart for a specified intermediate load and an arbitrary final load. Hinkle and Morse originally referred to the work of M. F. Spotts [3] in stating that such a spring should be designed for a final load equal to twice the intermediate load. C. W. Bert, a discusser to that paper, rightly pointed out the inconsistency of using Spotts' work, where dead coils are neglected, with their own work, which introduces the effect of dead coils. Mr. Bert's suggestion to use Jennings' [4] method of treating intermediate loads, however, is also in error. Jennings included the effect of dead coils, but did not include the effect of the Wahl correction factor. Since Hinkle and Morse included both the effect of dead coils and the Wahl correction factor, any method of handling intermediate loads must also be based on these two quantities. Hence the use of this design chart, with the minimum functions, is limited to problems where the final load is given.

References

3 M. F. Spotts, "Design of Helical and Leaf Springs for Minimum Weight," *Journal of Applied Mechanics*, vol. 20, TRANS. ASME, vol. 75, 1953, p. 435.

4 J. Jennings, "Springs of Minimum Weight," *Machinery*, vol. 52, 1938, pp. 707–709.



Errata to

Synthesis of Epicyclic Gear Trains Using the Velocity Ratio Spectrum[1]

By R. C. Brewer

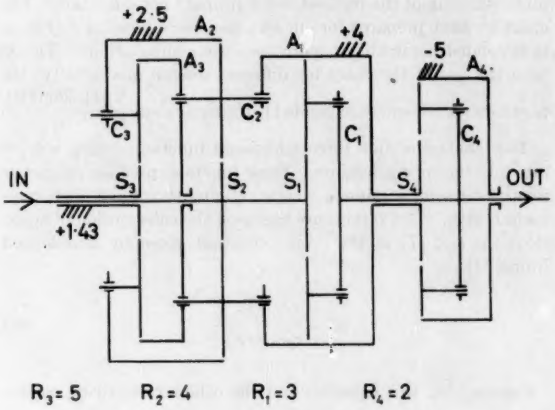
NOTE: Because of a misunderstanding in procedures at technical sessions, certain errors were inadvertently omitted from the published version of this paper.

These corrections follow:

page 177 (*a*) the penultimate line of the first column should read: "i.e., equation (1) with $R = R_2 = 4$

(*b*) the first three equations in the righthand column should likewise have the 5 changed to 4.

(*c*) in Fig. 9 sun wheel S_1 is shown connected to carrier C_1. The corrected Fig. 9 is reproduced here.



[1] Published in the August, 1960, issue of TRANS. ASME, Series B, JOURNAL OF ENGINEERING FOR INDUSTRY, pp. 173–178.

Journal of ENGINEERING FOR INDUSTRY

Information for Authors

• Manuscripts should be submitted in final form to the Editor. Each manuscript must be accompanied by a statement that it has not been published elsewhere nor has it been submitted for publication elsewhere. A paper which would occupy more than 8 pages of the Journal may be returned to the author for abridgment.

• The author should state his business connection, the title of his position, and his mailing address. A short abstract (50 to 100 words) should be included on the first page immediately preceding the introductory paragraph of the paper.

Three copies of the manuscript are required. One of these must be a carefully prepared printer's copy, typed in double spacing, on one side of the page only, with wide margins, on $8^1/_2$ by 11-in. opaque white paper. Mimeographed manuscripts, if prepared with exceptional care, can be accepted provided they are completely edited.

• As far as possible, all mathematical expressions should be typewritten. Greek letters and other symbols not available on the typewriter should be carefully inserted in ink. Care should be taken to distinguish between capital and lowercase letter, between zero (0) and the letter (*O*), between the numeral (1) and the letter (*l*), etc. A letter representing a vector cannot be printed with an arrow above or below it. The letter should be underscored with a single wavy line, wherever it appears in the text, to designate boldface type.

• A list of symbols carefully marked for the use of the editor (thus: κ, Greek l.c. kappa), if it has not been included in the body of the paper, should accompany the manuscript on a separate page.

• Before preparing a manuscript the author should study printed articles in the Journal, with special attention to the form and style of mathematical expressions, tables, footnotes, references, and abstract. Numbers that identify mathematical expressions should be enclosed in parentheses. Numbers that identify references at the end of the paper should be enclosed in brackets. Care should be taken to arrange all tables and mathematical expressions in such a way that they will fit into a single column when set in type. Equations that might extend beyond the width of one column (fractions that should not be broken or long expressions enclosed in parentheses) should be rephrased to go on two or more lines within column width. Fractional powers are preferred to root signs and should always be used in more elaborate formulas. The solidus should be used instead of the horizontal line for fractions wherever possible.

• A normal paper should not exceed 8 pages in the Journal. It is about 8000 words or 20 double-spaced typewritten pages exclusive of drawings. A Technical Brief should not exceed 750 words or about one page in the Journal inclusive of drawings.

• Originals and two copies of figures must accompany the manuscript. Line drawings should not be larger than $8^1/_2 \times 11$ in. and should be planned for reduction to column width. Lettering should be large enough to be clearly legible when the illustration is reduced. The originals of line drawings must be in India ink on white or pale blue tracing paper or tracing cloth. Photographs of equipment or test specimens must be glossy prints and should be used sparingly. Captions for figures should be typed double-spaced and included as the last page of the manuscript. The figure number and author's name should be written in the margin or on the back of each illustration.

• Titles of papers should be brief.

• Authors can obtain copies of the ASME Manual MS-4, "An ASME Paper," from the Editor and are urged to do so before drafting their papers in final form.

• Papers published in Journals of the Transactions of the ASME must be presented at a Meeting, either in person or by title. Scheduling of papers is in the hands of the Publications Committee.

• An author is entitled to 25 preprints free of charge (in the case of two authors, 15 each; three or more authors, 10 each). Larger quantities of preprints or reprints can be ordered from Editorial Department, The American Society of Mechanical Engineers, 29 West 39th Street, New York 18, N. Y. Quotations will be sent on request.

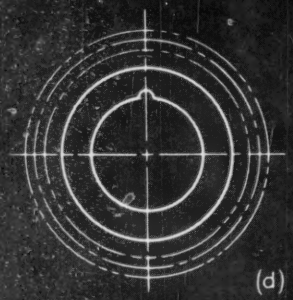
Please address all communications to:

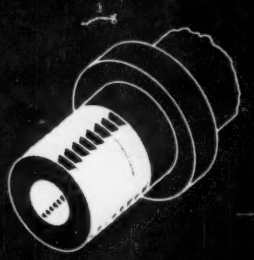
Editor, ASME,

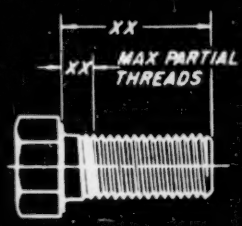
29 West 39th Street,

New York 18, N. Y.

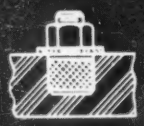
ASME Publications Committee







SUGGESTED SIZE OF CURL FOR INSIDE OR OUTSIDE DESIGNS



PREFERRED

NOT RECOMMENDED

Whether you need information on the fundamental principles of dimensioning, layout practices, pictorial sketching, projection, control and form tolerancing of geometric surfaces, the preparation of diagrams, or the special requirements for metal stampings, plastics, etc., you will find it in these twelve sections of the

AMERICAN DRAFTING STANDARDS MANUAL

SIZE AND FORMAT (Section 1), Y14.1 — 1957. $1.00

Deals with sheet sizes; the basic format; location of title and revision blocks, positioning of lists of material and drawing numbers, and print fold. Illustrated.

LINE CONVENTIONS, SECTIONING AND LETTERING (Section 2), Y14.2 — 1957. $1.50

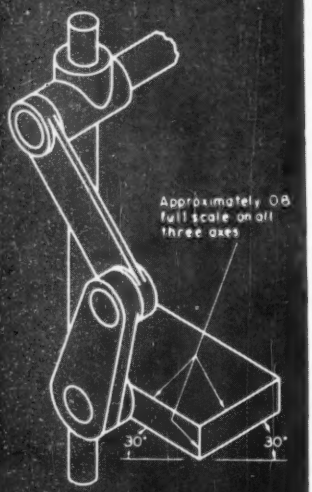
Scope Line symbols; visible and hidden lines; section, center and dimension lines; extension lines and leaders; cutting plane lines; break and phantom lines; section lining on detail and on assembly drawings; direction and spacing of general purpose section lines; the cutting plane; full and half sections; location of sectional views; lines behind the cutting plane; broken-out, revolved, removed, offset, auxiliary, and thin sections; sections through webs, shafts, bolts, pins; foresnortened projections and rotated features; intersections in section; sizes of lettering for different purposes.

PROJECTIONS (Section 3), Y14.3 — 1957. $1.50

Covering arrangement of views for multiple view orthographic projections, this section describes and illustrates practices in the choice and arrangement of the views; use of partial, alternate, removed and revolved views; auxiliary views; conventional breaks; and rounded and filleted intersections; developed views; descriptive geometry applications.

PICTORIAL DRAWING (Section 4), Y14.4 — 1957. $1.50

Here the various kinds of pictorial drawings are defined and the correct method of using them described. Axonometric, oblique, and perspective drawings are fully considered, and a number of examples presented showing the variety of positions in which the axes may be placed. Suggestions are given on the proper pictorial arrangement of sectional views, thread representation, the use of break lines, indicating fillets and rounds, unidirectional and pictorial plane dimensioning, shading, and phantom drawing.



DIMENSIONING AND NOTES (Section 5), Y14.5 — 1957. $2.00

Deals with the rules, principles, and methods used for specifying design requirements on drawings; illustrates how dimensions and notes should be used; and includes dimensioning practices for the control and form tolerancing of geometric surfaces.

SCREW THREADS (Section 6), Y14.6 — 1957. $1.50

Presents the approved methods of placing screw thread data on drawings and a considerable amount of information on thread tolerances useful to draftsmen. Typical drawing notes are shown along with specific practices for dimensioning.

GEARS, SPLINES, AND SERRATIONS (Section 7), Y14.7 — 1958. $1.50

Although this section is not intended to be a text book on gear design, it does give reasons for the methods shown and specified, so that the user will have some basic understanding of the need for more detailed gear-tooth dimensioning and specifications. Also included are the delineation and specification of Splines and Serrations.

FORGINGS (Section 9), Y14.9 — 1958. $1.50

In this standard is the needed information on preparing drawings for forgings, arranged in logical sequence and in usable form. The coverage includes properties of forgings, limitations, materials for forgings, production methods, parting line, forging plane, draft, fillets and corners, radii, web thickness, tolerances, forging drawings, including the parting line, unequal die depths, ribs, end draft for cylinders, intersections, and dimensioning.

METAL STAMPINGS (Section 10), Y14.10 — 1959. $1.50

Sets forth practices used by small parts manufacturers in the production of metal stampings as produced on standard types of punch presses. Here too, will be found help in laying out and delineating a product which will be subjected to metal stamping processes. References to methods of fabrication, die clearances, etc., are also included.

PLASTICS (Section 11), Y14.11 — 1958. $1.50

This section discusses basic preferred design and drafting practices specifically related to parts formed of plastic material. To assist the designer and draftsman in the delineation of plastic part drawings, a brief discussion of materials, manufacturing processes, and operations is included along with design and drawing hints for parts which are formed of the molding and laminating type plastics. Tolerance is also discussed from a general standpoint.

ELECTRICAL DIAGRAMS FOR ELECTRONICS AND COMMUNICATIONS (Section 15), Y14.15 — 1960. $1.50

Besides definitions and general information on electrical diagrams, this standard contains specific help and guidance in the preparation of single-line diagrams and schematic diagrams used by the electronics and communications industries. Items covered include layout; ground symbols; terminals; indication of parts; reference designations; methods of expressing resistance, capacitance, and inductance values; functional identification of parts; test points; additional circuit information; and single-line diagrams for microwave circuits.

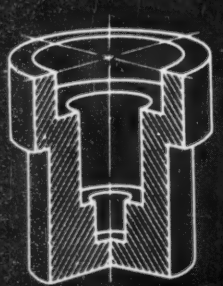
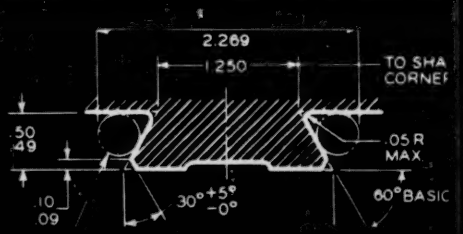
FLUID POWER DIAGRAMS (Section 17), Y14.17 — 1959. $1.50

This section not only establishes drafting practice for fluid power diagrams but also furnishes, for the first time, a comprehensive text and reference work on the subject. It explains in detail the data and notes which should accompany symbols and lines to make the diagram meet specific requirements. Types covered are: Pictorial diagram which may be used for quotation and piping the installation; cutaway diagram which is most often used for instruction purposes; graphical diagram which may be used for quotation, piping the installation, and for analyzing circuit operation; and a combination diagram which is used to emphasize purpose and operation of a portion of a system. Several examples of complete diagrams of each of these four types are included, also symbol data form ASA Y32.10.

Supplementing the recommendations are numerous sketches along with diagrams illustrating good drafting practices.

Price of these twelve sections and binder: $19.00 20% discount to ASME members.

THE AMERICAN SOCIETY OF MECHANICAL ENGINEERS, 29 W. 39th St., New York 18, N. Y.



Acceptable